Modelling Molecules for Shape-Driven Aggregation Three Case Studies

Doni, Giovanni

Awarding institution:
King's College London

Modelling Molecules for Shape-Driven Aggregation: Three Case Studies

by

Giovanni Doni

A Thesis submitted for the degree of

Doctor of Philosophy

in the

School of Natural & Mathematical Science

Department of Physics

March 2016

That's one giant step for a man...

Declaration of authorship

I, Giovanni Doni, declare that this Thesis, which I submit to King's College London in partial fulfilment of my application for the Degree of Doctor of Philosophy, gathers the original work completed under the supervision of Prof. Alessandro de Vita.

I certify that the work here presented was entirely carried on during my permanence as PhD candidate at King's College, from October 2011 to December 2014.

I declare that this thesis is based on my own original work. Where the contribution from collaborators is used, the extent and the quality of their contribution is stated and acknowledged.

All the work presented herein is original and was not previously submitted as an application for a degree or other qualification to any other institution.

Part of the work presented was published or submitted for publication:

- G. Doni, M.D. Nkoua Ngavouka, A. Barducci, P. Parisse, A. De Vita, G. Scoles, L. Casalis, G. M. Pavan, *Structural and energetic basis for hybridization limits in high-density DNA monolayers,* Nanoscale, 2013,5, 9988-9993.
- M. Riello, G. Doni, S. Filip, G. Martin, A. De Vita, *Pressure-induced conformation transition of o-phenylene solvated in bulk hydrocarbons,* Journal of Physical Chemistry, 2014,47, 13689-13696.

Giovanni Doni

Acknowledgments

I wonder what is left of a PhD when it finally comes to its end. Clearly, there is much more than the hundreds of pages which follow. There is the awareness and incredible personal development. There is the memory of some exciting as well as difficult moments. There is the path that I left behind during a long and difficult journey, a path that I shared with many bright people to which I am deeply grateful: it has been for me a privilege to have spent three years here at King's College, an enriching experience in the heart of London that I have undertaken with enthusiasm. There is, of course, the research that I carried out. This, and all of the above, would not have been possible without the support of the wonderful people that I met during this time.

In this respect, I would like to thank my supervisor, Alessandro De Vita, who offered me this valuable opportunity of growth, both from an academic and personal point of view. I am deeply grateful to Giovanni, an inspiring friend and colleague with whom I had the honour to collaborate with over the years. A mention goes to Erlend for his support on setting up some calculations and for the nice discussions we had together. A big thank goes to the fantastic group of people with whom I had the occasion to work, collaborate or simply to share ideas and opinions with on basically every possible topic: thanks to Giovanni, Gianmarco, Marco, Federico. I would also like to thank all the staff that contributed to making the Physics department at King's College a unique place.

I would like to wish all those people not explicitly mentioned herein, and to all the staff and future PhD students, a happy and fruitful voyage.

More than three years of hard work saw me establishing durable friendships. I would like to thank my dear friends and colleagues Paolo, Massimo, and Silvia: may these years last in our memory. I would also like to thank my flatmates with which I shared my life in London: Marco, Jutta, Claudia, Michele and Silvia.

I dedicate this thesis to my family and to my dearest friends that area carrying on with their studies and careers between Trieste, Berlin and Pisa: Simone, Martino, Teresa e Chiara, will catch up soon, I promise.

Last, I would like to thank Catherine for making London more of an house that a urban jungle: all of this would make much less sense to me without you.

Abstract

The exploitation of self-assembling for the development of reliable technological platforms depends on the characterisation of the interactions between molecules. The possibility of tuning these interactions in order to create complex aggregates bearing specific chemical/physical properties at the macroscopic level, is of crucial importance for the development of nanotechnology. In this framework, atomistic simulations constitute a powerful tool capable of providing fundamental contributions. Following this line, the present work combines three distinct studies carried out on as many molecular systems, where the use of a common set of computational techniques, based on classical potentials and Density Functional Theory (DFT), was addressed to understand the aggregation mechanisms displayed by poly-aromatic gelators, poly-aromatic hydrocarbons and nucleic acids. This thesis is structured as follows.

In the first chapter, a review of the main computational techniques used throughout this study is presented. The foundations of Molecular Dynamics (MD) are introduced first, with particular attention to the atomistic classical potentials and the main algorithms used for the integration of the equation of motions, the generation of the correct ensemble and for the solvation energies. An overview of DFT is also presented, completing the overview of the investigation methods adopted.

In the second chapter, the results of the research carried out to characterise the interactions between a particular type of poly-aromatic gelators, is presented. In particular, the study focuses on the individuation of a set of binding configurations able to provide a rationale for the sensitivity of gel formation to the enantiopurity of the solution.

The third chapter focuses on the work conducted in collaboration with the British Petroleum. Over the course of three years, in the framework of two consecutive research projects, the design and development of fuel additives was supported by a series of computational studies aimed at assessing the effectiveness of different molecular structures and functional groups for the interaction with aromatic substrates. Here, the results of DFT calculations carried out on a series of polycyclic aromatic compound are presented, together with a correction scheme

applied to a widely used classical potential for $\pi - \pi$ interactions; a model dendritic structure with multiple phenyl terminations is proposed and tested with a molecular model, demonstrating the effectiveness of a multivalent structure for the binding with highly defective polyaromatic substrates.

In the fourth chapter, the work conducted for the characterisation of DNA Self-Assembled monolayer is presented. MD simulations of DNA high-density monolayers of increasing hybridisation states have been carried out for the characterisation of the energetics of the nanostructure in the framework of surface hybridisation; this allowed to highlight the intrinsic hybridisation limit reported by several independent experimental works.

In the conclusive section, the results presented in the previous three chapters are summarised, highlighting the major achievements and element of novelty of the research. The critical review of the work is combined with some considerations on the expected developments and future experimental outcomes.

Contents

1 Computational methods **12**

1.1 Introduction . . . 12

1.2 Molecular Dynamics . . . 14

1.2.1 The Force Field . . . 14

1.2.2 Integration of the equation of motion . . . 18

1.2.3 Periodic boundary conditions and long-range corrections . . . 19

1.2.4 Statistical ensembles . . . 21

1.2.4.1 Langevin thermostat . . . 22

1.2.4.2 Nosé-Hoover . . . 23

1.2.4.3 Barostats . . . 24

1.3 Solvation models and electrostatics . . . 26

1.3.1 Explicit Solvation . . . 26

1.3.2 Implicit Solvation . . . 27

1.4 Density functional theory . . . 31

1.4.1 The basis set . . . 33

1.4.2 Pseudopotentials . . . 34

1.4.3 Force evaluation . . . 35

1.5 Minimisation algorithms . . . 36

1.6 Softwares . . . 37

2 Chiral dependent gelation of poly-aromatic LMOG **38**

2.1 The gelation of binaphtol derivatives . . . 38

2.1.1 The binaphtol derivatives . . . 40

2.2 The experimental findings . . . 43

2.2.1 Analysis of NMR measurements . . . 43

2.2.2 Gel micrographies . . . 45

2.3 Ultrasounds in chemistry . . . 49

2.4 Preliminary computational studies . . . 52

2.4.1 The two conformers . . . 54
2.4.2 The transition energy barrier . . . 55
2.5 Solvation energies for different solvents . . . 57
2.6 Binding of binaphtol molecules . . . 59
2.7 The dimers . . . 62
2.8 10-dimers clusters . . . 66
2.9 Sampling the space of rotations . . . 69
2.9.1 Uniform sampling . . . 70
2.10 Gas-phase cohesive energy on a fcc lattice . . . 73
2.11 Sampling the dimer-dimer interaction energy . . . 76
2.11.1 The Monte Carlo sampling . . . 77
2.11.2 Binding configurations of dimer A . . . 78
2.12 Ising-like model for aggregation . . . 86
2.12.1 Results . . . 89
2.13 Summary . . . 92

3 $\pi - \pi$ stacking for polyaromatic substrates 95
3.1 Introduction . . . 96
3.1.1 Polycyclic Aromatic Hydrocarbons . . . 97
3.2 Binding energy for PAHs . . . 99
3.2.1 Computational details . . . 99
3.2.2 Calculated energies . . . 100
3.2.3 DFT-based correction to the classical potential for $\pi - \pi$ interactions . . . 102
3.3 Benzene-terminated dendron . . . 107
3.3.1 Model the substrate/solvent interface . . . 109
3.3.2 The hydrocarbon model . . . 110
3.3.3 The sticking mechanism . . . 113
3.3.4 Adlayer cohesion energies . . . 116
3.4 Summary . . . 118

4 Hybridisation limit for high-density DNA monolayers 120
4.1 Introduction . . . 120
4.1.1 DNA Self-Assembled Monolayer . . . 121
4.1.2 Surface hybridisation of DNA monolayers . . . 123
4.1.3 Hybridisation limit of high-density DNA monolayers . . . 125
4.1.4 Computational contributions . . . 126

4.2 HDM model and computational details 128
4.2.1 Modelling the DNA monolayer 128
4.2.2 Parametrisation and simulation procedure 130
4.2.3 Experimental setup 132
4.3 Energy and structural analysis 133
4.3.1 HDM energy profiles 133
4.3.2 HDM Structural analysis 136
4.4 Coarse-grained model of HDMs 141
4.5 Summary 142

5 Outlook and conclusions **144**

Bibliography **148**

List of Figures

2.1 Vials containing the gel . . . 40
2.2 Reaction mechanism - the two conformers . . . 41
2.3 NMR solvent fractions . . . 45
2.4 SEM micrographies of the aggregates . . . 47
2.5 SEM micrographies of the fibres - 1 . . . 47
2.6 SEM micrographies of the fibres - 2 . . . 48
2.7 TEM micrographies . . . 48
2.8 SEM micrographies of incomplete gel formation . . . 49
2.9 Relaxed structures for DFT and classical potential . . . 55
2.10 Energy profiles from NEB calculations . . . 56
2.11 Helix aggregate . . . 60
2.12 Most relevant binding configurations . . . 61
2.13 Interlocking dimers . . . 63
2.14 RR-SS and RR-RR methylated dimers . . . 64
2.15 NMR solvent fractions model fit . . . 67
2.16 10-dimers clusters average energies . . . 68
2.17 Angle representation . . . 70
2.18 Sampling of rotations . . . 73
2.19 Tetrahedral lattice of fcc clusters . . . 74
2.20 Dimer-dimer binding energy distribution . . . 79
2.21 Energy distribution for minima . . . 80
2.22 Most relevant dimer-dimer binding configurations . . . 81
2.23 Dimer-dimer binding energy minima . . . 83
2.24 Details of two binding configurations between methylated dimers . . . 85
2.25 Repetitive binding for dimers . . . 86
2.26 Lattice model for aggregation . . . 88
2.27 Lattice equilibrium average energy . . . 90
2.28 Equilibrated lattice configurations . . . 91

2.29 Specific heat map 92

3.1 Benzene dimers configuration of minimum energies 98
3.2 PAH molecules 100
3.3 Normalised binding energies for PAH dimers and PAH at graphene 103
3.4 12-6 LJ vs DFT energy profiles 105
3.5 9-6 LJ vs DFT energy profiles 106
3.6 Model dendron molecule 108
3.7 Dendron physisorption at graphene 109
3.8 Model surfaces 111
3.9 Reference diesel fuel composition 111
3.10 Distance from graphene of the phenyl groups 113
3.11 Dendron anchoring mechanism 114
3.12 Sticking mechanism 115
3.13 Surface ordering of hexadecane molecules 116

4.1 Nano-shaving technique 122
4.2 Functionalisation through surface hybridisation of monolayers 124
4.3 DNA sequence 128
4.4 HDM model system set up 129
4.5 HDM model systems for different hybridisation states 130
4.6 Average per strand energy vs hybridisation fraction 134
4.7 AFM height measurements - surface renderings 137
4.8 Calculated monolayers' heights 139
4.9 Atomic density profile $\rho(z)$ 140

Chapter 1

Computational methods

In this first chapter, an overview of the computational methods used throughout this study is presented. The fundamental principles of classical Molecular Dynamics are introduced first, together with a review of the main algorithms for the integration of the equation of motion, the sampling of the different statistical ensembles and a review of the most used solvation models. A brief overview of the foundation of DFT is also presented. For a comprehensive treatment of all the topics covered herein, extensive readings can be found in [1, 2, 3].

1.1 Introduction

Thanks to the advancements in electronics and the birth of computers, a set of problems that could not easily be solved analytically in statistical mechanics, started to be tackled with the aid of numerical techniques. In this framework, one of the methods that was soon developed was *Molecular Dynamics* (MD). This term classifies a set of computational techniques progressively refined over the course of the last fifty years used to calculate average physical quantities of atomistic systems aiming at providing insight into experimental results.

The idea behind classical MD is fairly straightforward. An atomistic system, being either a group of molecules or a crystal, is represented through a set of point particles through their coordinates. An interatomic potential defines the forces acting between atoms that upon evaluation, are inserted into the equations of motion. The iterative solution of the equations of motion at each time-step gives a set of new positions and velocities that constitute the trajectory of the system over the integration time. The equilibrium trajectory samples a statistical ensemble and can be used to calculate physical quantities thanks to the ergodic principle.

One of the first applications of MD was accomplished by Alder et al. [4] in the end of 1950s: a hard sphere model for non-interacting particles was successfully implemented for the calculation of the equation of state, despite the limitations of early computational resources. A decade later, the simultaneous integration of coupled differential equations became possible, and the simulation of the first Lennard-Jones models [5] was successfully carried out. After these initial works, MD developed rapidly. The aim of modelling real liquids led in 1974 to the first simulation of liquid water [6]. The attention was soon driven to bigger molecules and in 1977 the first protein was studied through MD [7]. Also solid-state physics soon benefitted from the advancement of MD: the fitting of the classical potential for metals led to calculate the elastic properties of crystals and the correct reproduction of the vibrational spectra; phases transitions, diffusion of defects and the influence of substituent atoms have been historically successfully studied.

Over the course of the last two decades, new techniques and algorithms have been developed to deal with more complex and larger systems, expanding the capabilities of MD. On one side, the research has focused on developing methods able to extend the precision of classical MD by embedding Quantum Mechanics corrections within classical potentials: on a smaller portion of the system the forces are calculated with *ab initio* methods, while a corrective term is added globally to the force field to insure continuity. The increased accuracy of this type of hybrid techniques allow to account for bond breaking events and chemical reactions, that are fundamental in the study of a broad series of local phenomena, such as enzymatic reactions and fracture in solids [8].

Another set of problems limiting the applicability of MD has to deal with the extent of phase space that can be accessed in feasible simulation time. Large systems constituted by many atoms are, in fact, computational demanding and require long integration times to gather a sufficient statistics. Parallel executions of calculations on several replicas of the same system at different conditions i.e. different temperatures or using different Hamiltonians, together with the introduction of biasing potential methods, such as in metadynamics or umbrella sampling, can greatly improve the sampling of the free energy surface, overcoming energy barriers otherwise limiting the diffusivity in the phase space.

Together with the refinement of more efficient computational methods, the expanding growth of MD has been fuelled by the rapid increase in computing power: not only the performances of processors have exponentially increased but also more efficient architectures and programming strategies have been developed. This has progressively enabled larger systems to be simulated and for longer simulation times. While the first applications of MD of solvated proteins, nucleic acids and polymers consisted of systems of few thousands of atoms and integration times of a few nanoseconds, nowadays, microsecond-long simulations of systems having a million of particles are

possible. A whole set of systems are, indeed, studied through MD, aiming to answer questions arising from various fields of research, including biology. In particular, a great deal of attention is devoted to study of protein folding, to model the mechanical properties of membranes [9, 10] and the permeation of molecules through them [11], to discover the mechanism leading to bio self-assembling [12], to unravel the functioning of protein complexes on a molecular level, such as membrane receptors and ion channels [13].

The success and the diffusion of MD since its beginnings is due to its unique contribution to the research bridging the gap between theory and experimental results. MD demonstrated to be particularly effective in:

- studying the relationship between molecular interactions and the macroscopic behaviour of system;
- sampling the conformational space of complex molecular system;
- estimating the free energy differences involved in the complexation mechanisms, ubiquitous in molecular recognition processes;
- highlighting atomistic details that cannot be easily captured by experiments.

The importance of MD was recently sealed by the international scientific community with the assignment of the 2013 Nobel Prize in Chemistry to three of the major contributors in the field, M. Karplus, M. Levitt and A. Warshel, "for the development of multi-scale models for complex chemical systems".

Like any other models, MD is based on some approximations that have to be consistent with the underlying physics. The reliability of the model is fundamental for producing meaningful and consistent results and depends on the correct parametrisation of the interactions and the stability of the integration and thermostatting algorithms. These topics are covered in the following sections.

1.2 Molecular Dynamics

1.2.1 The Force Field

Whenever the electronic structure of an atomistic system is fundamental for the determination of its properties it is necessary to rely on the level of detail of ab initio calculations. Processes such as chemical reactions and charge transfer can be essentially modelled only with the help

of quantum mechanical methods (e.g. DFT, Tight Binding etc.). Unfortunately, the use of these techniques comes at the expense of high computational requirements that severely limits the size of the system and the simulation times that can be accessed, which might be not sufficient to characterise the mechanism under investigation. However, there is a broad series of cases for which the use of classical potential can substitute quantum mechanical approaches without significantly compromising the accuracy of results. This is, for example, the case of those mechanisms that depends on long-range, non-bonded interactions between molecules.

Since the masses of electrons and nuclei differ by three orders of magnitude, the *Born-Oppenheimer* approximation assumes their motion is decoupled and thus the wavefunction of the system can be factorised into an electronic $\psi(\mathbf{r};\mathbf{R})$ and nuclear part $\chi(\mathbf{R})$. As a consequence, the Schrodinger equation can be split into two parts as

$$[T_e + V_{tot}(\mathbf{r},\mathbf{R})]\ \psi(\mathbf{r};\mathbf{R}) = \epsilon(\mathbf{R})\ \psi(\mathbf{r};\mathbf{R}), \tag{1.1}$$

$$[T_n + \epsilon(\mathbf{R})]\ \chi(\mathbf{R}) = E\ \chi(\mathbf{R}), \tag{1.2}$$

where T_e and T_n represent the kinetic operator for electrons and nuclei, V_{tot} is the total potential for the electrons and E is the energy of the system. $\epsilon(\mathbf{R})$ is the value of the ground state energy found by solving the electronic part of the Hamiltonian, as a function of the fixed position of nuclei $\mathbf{R}$. Effectively, $\epsilon(\mathbf{R})$ can be seen as the potential energy, mapping the contribution of chemical bonds between atoms, as well as non-bonded interactions. The forces acting on the nuclei can be calculated by considering the electrons as being at all times at their ground state.

If $\epsilon(\mathbf{R})$ was known beforehand, there would be no need for the equation 1.1, while the equation for the nuclei would be reduced to a entirely classical Hamiltonian. Classical *force fields* (*ff*) were developed precisely for this reason, with the aim of mapping $\epsilon(\mathbf{R})$ as the sum of simple terms having various analytical forms, accounting for all the possible type of interactions between atoms. A general expression of $V(\mathbf{r}^N)$ can be expanded in the sum of several components,

$$V(\mathbf{r}^N) = \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2(\mathbf{r}_i,\mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j} v_3(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) + \ldots, \tag{1.3}$$

where each v_i depends on one or more particles' positions. The precise analytical form of 1.3 depends on the specific system under investigation and accounts for different terms having a defined physical meaning. Normally, the parameters can be either derived by fitting ab initio calculation energy profiles or, alternatively, can be refined to match experimental results.

For a molecular system, the general expression of a *ff* combines the contributions of different

terms modelling intra- and inter-molecular interactions and can be written as follows,

$$V(\mathbf{r}^N) = \sum_{bonds} \frac{K_b}{2}[b - b_0]^2 + \sum_{angles} \frac{K_\theta}{2}[\theta - \theta_0]^2 + \sum_{dihedrals} \frac{K_\phi}{2}[1 + cos(n\phi - \delta)] \qquad (1.4)$$
$$+ \sum_{i,j} 4\epsilon_{ij}\Big[\big(\frac{\sigma_{ij}}{r_{ij}}\big)^{12} - \big(\frac{\sigma_{ij}}{r_{ij}}\big)^{6}\Big] + \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon r_{ij}}.$$

The first three terms represent the contributions originating from the decomposition of bond energy into two, three and four body terms. The first one (*two-body*) represents the covalent bond stretching interaction and is often approximated with a harmonic potential around a minimum bond length of b_0 and force constant equal to K_b. The second term (*three-body*) is the bond bending energy, and is again expressed with a harmonic potential applied to the angle formed by two consecutive bonds, being θ_0 and K_θ the equilibrium angle and force constant, respectively. The third term (*four-body*) adds the contributions of torsion angles between four consecutive atoms and it is modelled by a periodic cosine function of periodicity n, where δ and K_ϕ are the equilibrium torsion angle and force constant.

The remaining part of eq. 1.4 contains the *non-bonded* terms. This part of the potential accounts for the various types of interactions that are established between atoms, which are not forming a covalent bond, and models dispersion forces and short range electrostatics. Different analytical forms can be found for different *ff*, however pair additive potentials are the most frequently used. A 6-12 Lennard-Jones (LJ) potential is commonly adopted to represent the van der Waals interactions, while a Coulomb term accounts for the contribution arising from the partial charges q_i assigned to each atom. The coefficients, ϵ_{ij} and σ_{ij}, are specific for each atom pair and are calculated from the single atoms' parameters ϵ_{ii} and σ_{ii} through the Lorentz-Berthelot mixing rule [1]. The potential energy of the system is thus obtained by summing all the previous terms over all the bonds, angles and torsions for each molecule, and over all pairs.

As mentioned earlier, the one reported above is not the unique expression for a *ff* but various functional forms can be chosen with a set of corresponding parameters, in order to model some particular types of interaction not precisely mapped by the terms reported in equation (1.4). Covalent bonds, for example, require sometimes to be modelled by a Morse potential to account for bond breaking events; a more complex analytical form of the angle and torsion terms can be necessary to represent adequately bond bending and the steric hindrance arising from the rotation of particular chemical groups; moreover, dipole-dipole contribution can be added to account for atom polarisation.

Force fields have been developed and optimised over the course of the last three decades to

[1] The LJ parameter for a pair of atoms i, j is obtained averaging the corresponding single atoms values ϵ_i and σ_i as follows: $\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j)$ and $\epsilon_{ij} = (\epsilon_i \epsilon_j)^{1/2}$.

deal with the majority of the systems of interest. This continuous work of parametrisation and refinements aims at improving the existing *ff*s by testing their performances against the experimental results. The most famous and widespread *ff* available for molecular system are *GAFF* [14], *CHARMM* [15], *GROMOS [16]* and *OPLS* [17]. Each of these names identifies a group of databases of parameters characterised by a similar analytical expression for the potential. Originally implemented during the '80 in different research groups along with the first MD codes, *ffs* are available in multiple versions, each of which is either an updated or a modification of the original implementation, and are optimised to increase the accuracy of the results for a precise class of molecules, for a given environment and thermodynamic conditions. Historically, proteins and nucleic acids are among the molecules that first started to be investigated first through MD [18]: as a consequences, a huge effort was devoted to the development of *ff* specifically targeted to study this kind of biological systems. In addition, the interest of applied chemistry for polymers, lipids, solvents and other types of common organic and inorganic compounds fuelled the proliferation of *ff*, targeted to a specific research purpose. Force fields can thus be classified along different lines:

- the type of system e.g. hydrocarbons, lipids, nucleic acids;
- the type of environment e.g. liquid phase, solid phase, polar or non-polar solvent;
- the range of thermodynamic conditions e.g temperature and pressure ranges;
- the type of interaction e.g. bond stretching, VdW, dipoles;
- the analytical form of the *ff* e.g. harmonic potential, Lennard-Jones;

The heterogeneity of systems that are commonly studied with MD and the abundance of *ff* indicate that there is not a unique choice for the parametrisation of a system. Classical potentials are, in fact, optimised to closely reproduce a set properties, while other quantities might not be estimated with the same accuracy; no *ff* is, in fact, able to provide absolutely reliable results and each *ff* is expected to perform differently. It is, therefore, necessary to verify the capability of a given *ff* to produce accurate results through the comparison with alternative parametrisations and the validation with experimental data. A trade off between accuracy and transferability is, in most cases, a desirable target.

1.2.2 Integration of the equation of motion

The Hamiltonian of the system of positions r^N is straightforwardly determined as the sum of the kinetic and the potential energy,

$$H = \sum_i \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}^N), \tag{1.5}$$

where p_i indicates the momentum of the i-th particle and $V(\mathbf{r}^N)$ is the potential energy expressed by the *ff*. The force acting on each atom can be calculated upon differentiation with respect to the respective coordinate. If the potential is pair additive, though, the force f_i is obtained as the sum over atom pairs of the non-bonded terms plus the forces due to bonded interactions (which may depend on four atoms at most),

$$\mathbf{f}_i = \sum_j \mathbf{f}_{ij}. \tag{1.6}$$

The forces acting on each particle i can be inserted in the corresponding Newton's equation of motion, which are integrated using numerical techniques based on finite difference schemes. Time is discretised in intervals of length $\triangle t$; by knowing the positions and its time derivatives (up to a certain order) at the time t, integration algorithms generate the new positions and velocities at the subsequent times $t + n\triangle t$. The integration schemes that are normally applied in MD are based on a Taylor expansion of the positions at $t + \triangle t$ and $t - \triangle t$, whose difference yields the following expression,

$$\mathbf{r}(t + \triangle t) + \mathbf{r}(t - \triangle t) = 2\mathbf{r}(t) + \ddot{\mathbf{r}}(t)\triangle t^2 + \mathcal{O}(\triangle t^4). \tag{1.7}$$

The updated position at the time $t + \triangle t$, can then straightforwardly calculated as

$$\mathbf{r}(t + \triangle t) \simeq 2\mathbf{r}(t) - \mathbf{r}(t - \triangle t) + \frac{\mathbf{f}_i(t)}{m_i}(t)\triangle t^2. \tag{1.8}$$

This simple scheme, known as the Verlet algorithm [19], combines the advantages of numerical stability, time-reversibility and energy conservation. However, it does not explicitly include the velocities, that have to be calculated by a central difference, as

$$\dot{\mathbf{r}}(t) = \frac{\mathbf{r}(+\triangle t) - \mathbf{r}(t - \triangle t)}{2\triangle t} + \mathcal{O}(\triangle t^2). \tag{1.9}$$

The calculation of positions and velocities with two different orders of precision (fourth and second order, respectively) is not desirable since it limits the size of the time-step for an accurate

estimation of velocities. To solve this problem, various higher-order integration schemes have been developed, among which the two most common integrators used in MD are the *Leap-frog* and the *Velocity-Verlet* algorithm. In the leap-frog algorithm, the velocities are calculated on a staggered grid at the mid-steps $t \pm \frac{\triangle t}{2}$, as

$$\mathbf{r}(t + \triangle t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t + \frac{\triangle t}{2})\triangle t, \tag{1.10}$$

$$\dot{\mathbf{r}}(t + \frac{\Delta t}{2}) = \dot{\mathbf{r}}(t - \frac{\Delta t}{2}) + \frac{\mathbf{f}_i(t)}{m_i}\triangle t. \tag{1.11}$$

Eq. 1.11 is carried out first, yielding the velocities $\dot{\mathbf{r}}(t + \frac{\Delta t}{2})$, from which the position at the following time-step are derived through eq. 1.10. The velocity at the time t are not made available directly, but can be calculated through the linear interpolation of $\dot{\mathbf{r}}(t - \frac{\Delta t}{2})$ and $\dot{\mathbf{r}}(t + \frac{\Delta t}{2})$. In the Velocity-Verlet algorithm, instead, positions and velocities at time t are used to integrate the equations of motion; the velocities at the previous half step are not required,

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\Delta t + \frac{\mathbf{f}_i}{2m_i}\triangle t^2, \tag{1.12}$$

$$\dot{\mathbf{r}}(t + \Delta t) = \dot{\mathbf{r}}(t) + \frac{\mathbf{f}_i(t) + \mathbf{f}_i(t + \triangle t)}{2m_i}\triangle t. \tag{1.13}$$

Both the two methods ensure an accurate calculation of the trajectories and a small energy drift. Higher order algorithms can further increase the precision at the expenses of a more demanding computational load. For this reason, predictor and corrector methods are less commonly used, unless more accurate integration algorithms are needed in conjunction with specific type of calculations such as for hybrid QM/MM approaches.

1.2.3 Periodic boundary conditions and long-range corrections

The introduction of periodic boundary conditions is necessary to model a bulk atomistic system. Particles are confined within a box (usually orthorhombic) of finite dimensions $\mathbf{L}$ and the box is virtually replicated in all directions. Clearly, this implies that the sum in eq. 1.6 needs to include the contribution of neighbouring particles belonging to the adjacent replicas. As a consequence, the calculation of forces should be performed over an infinite number of pairs; for the electrostatic energy term this means that the sum is written as

$$U_{coul} = \frac{1}{8\pi\varepsilon}\sum_{i,j}^{N}\sum_{n\epsilon Z^3}\frac{q_i q_j}{|\mathbf{r}_{ij} + n\mathbf{L}|}, \tag{1.14}$$

where n is and integer accounting for periodic images. Since the contribution of short-range potentials decays rapidly with the distance, the *minimum image convention* is applied, limiting the force evaluation to the nearest images of the N-1 atoms. For systems of thousands of atoms though, this can still result in an unreasonably large number of force evaluations at each time step. To circumvent this problem, a spherical cutoff r_c is used, setting the pair potential to zero for distance exceeding r_c; effectively, this is implemented by just expanding the force evaluation to those pair lying within the volume defined by r_c. To ease the search for pairs within this volume, a list of neighbours is created and updated with a certain frequency.

The use of cut-offs clearly introduces small errors that can be limited by the choice of a sufficiently large cutoff radius, normally in the order of 10 Å; this, in agreement with the minimum image convention, should not be greater that half of the smallest box dimension. Long-range corrections are thus often necessary to compensate the errors introduced by the truncation of the potential, in particular, for the electrostatic interactions between point charges, for which the energy decays with the inverse power of the separation distance. Various schemes have been developed to deal with these corrections, most of which originates from the Ewald summation method [20] developed for periodic crystals. The idea relies on breaking down the slowly converging sum of eq. 1.14 into two quickly converging terms in the real and the reciprocal space plus a constant[2],

$$U_{coul} = U^{(r)} + U^{(k)} + U^{(s)}. \tag{1.15}$$

In eq. 1.15, $U^{(r)}$ is the real space sum described earlier; $U^{(k)}$ is the sum of performed over the reciprocal space; $U^{(s)}$ is the contribution arising from the self-energy, which vanishes with the condition of charge neutrality of the whole system. The previous terms can then be written explicitly as

$$U^{(r)} = \frac{\alpha}{8\pi\varepsilon_0} \sum_{i,j}^{N} \sum_{n\epsilon Z} q_i q_j \frac{erfc(\alpha \mathbf{r}_{ij})}{|\mathbf{r}_{ij} + n\mathbf{L}|}, \tag{1.16}$$

$$U^{(k)} = \frac{\alpha}{2\varepsilon_0 L^3} \sum_{k\epsilon R^3} \frac{2\pi}{k^2} \exp(-\frac{k^2}{4\alpha^2}) \sum_{i,j} q_i q_j \exp(-i\mathbf{k}\cdot\mathbf{r}_{ij}), \tag{1.17}$$

$$U^{(s)} = -\frac{\alpha}{\pi^{\frac{1}{2}}} \sum_{i} q_i^2. \tag{1.18}$$

The inner sum in eq. 1.17 is the Fourier transform of the charge density, while α is a parameter tuning the relative weight of the real and the reciprocal space contribution. The sum in reciprocal space is performed over a set of vectors through the introduction of an additional cut-off in k-space. The computational cost of Ewald summation scales with $N^{\frac{3}{2}}$ (for an appropriate choice

[2] Additional contributions to eq. 1.15 can include long-range corrections for dipoles and dispersion forces.

of α [21]) and it is not considered convenient for large systems. To improve the performances of the reciprocal sum, other methods are often preferred. *Particle Mech Ewald* (PME) [22] and *Particle-Particle Particle-Mesh* (PPPM) [23] are two of the most efficient and widely used: by exploiting a reduced representation of charges onto a mesh, the charge density is interpolated allowing for a faster convergence of Discrete Fourier Transform; in this way, the computational cost is decreased to NlogN.

1.2.4 Statistical ensembles

The importance of MD relies on the fact that the trajectories generated by the integration of the equations of motion can be used for the calculation of the properties of an atomistic system. According to statistical mechanics, in fact, physical observables correspond to averages performed over a statistical ensemble, which is determined by a certain distribution of the phase space variables. Importantly, the *ergodic principle* states that the ensemble averages are equivalent to time averages, in the limit of sufficiently long times. The implication of this for MD is that, given a certain properties A, which is a function of positions and momenta,

$$A(r^N, p^N) = f(\mathbf{r}_i(t), \ldots, \mathbf{r}_N(t), \mathbf{p}_1(t), \ldots, \mathbf{p}_{n(t)}), \tag{1.19}$$

its average can be approximated by a time average over a subset of the trajectory, at the times t_i as

$$\langle A \rangle = \frac{1}{Z} \int A\rho(\mathbf{r}^N, \mathbf{p}^N) d\mathbf{r}^N d\mathbf{r}^N = \lim_{t_n \to \infty} \frac{1}{t_n} \int_0^{t_n} A(t)dt \approx \frac{1}{N_T} \sum_i^{N_T} A(t_i), \tag{1.20}$$

where ρ is the probability distribution and Z is the partition function. The ensemble average corresponds to the time average for an adequately large number of samples N_T and at the condition that the that samples are not correlated, i.e., $t_{i+1} - t_i$ is higher than the correlation time. More importantly, these averages require that the trajectories are drawn from the appropriate ensemble. For a fixed number of particle N and a fixed volume V, the total energy of the system is conserved and the trajectories generated by the solution of the Newton's equation of motion sample the *micro-canonical ensemble*, also referred as *NVE*. In many cases, however, it is more convenient to perform the calculation in the *canonical ensemble (NVT)*, in which it is temperature, rather that the total energy, to be kept constant. For a very large system of many thousands of atoms, the thermodynamic limit is approached and the system acts as its own heat bath: at this condition, the *NVE* and *NVT* ensembles become equivalent. For smaller systems, though, a thermostat is needed to keep the temperature constant by exchanging heat with a fictitious bath. Many different thermostats have been proposed: the Andersen

thermostat [24] consists of periodically setting the velocity of a fraction of randomly selected particles to the values extracted from correct Maxwell distribution; Berendsen [25] implemented an agile mechanism of proportional scaling of the momenta to match the reference temperature. The simplicity and the stability of the these two methods is counterbalanced by a non-rigorous sampling of the NVT ensemble. For this reason, the *Langevin* and the *Nosé-Hoover* thermostats are often preferred.

1.2.4.1 Langevin thermostat

The Langevin thermostat [26] is a simple thermostatting algorithm that samples ergodically the canonical ensemble by coupling the system's degrees of freedom to an external bath of harmonic oscillators. The interaction with the external bath's particles reduces to a dissipative term proportional to the velocities, while a random force represents the effect of the elastic collisions with the fictitious particles. The deterministic equations of motion for the extended system are transformed into a stochastic form, known as Langevin equations:

$$m_i\ddot{\mathbf{r}}_i = -m_i\gamma\dot{\mathbf{r}}_i + f_i(t) + \sqrt{2\gamma m_i kT}w_i. \tag{1.21}$$

The random forces w_i are drawn from a Gaussian distribution with zero mean and a variance that is scaled accordingly to the velocity autocorrelation function through the Stokes-Einstein fluctuation-dissipation relation for the diffusion coefficient as follows:

$$\langle\dot{\mathbf{r}}_i(t)\dot{\mathbf{r}}_i(t+s)\rangle = 2m_i\gamma kT\delta(s)\delta_{ij}, \tag{1.22}$$

where m_i is the particle mass and T the target simulation temperature, i.e., the temperature of the heat bath. The damping coefficient γ has the dimension of $1/time$ and represents the characteristic time scale over which the thermostat acts. The performances of the thermostat depend on the choice of γ: for large values, the stochastic term prevails, resulting in a strongly stochastic thermostat that recovers a Monte Carlo sampling; conversely, for smaller values, a more deterministic thermostat is recovered, at the expenses of a restriction of the phase space explored by the simulation. For liquid phases, γ is routinely chosen using the water diffusion coefficient as reference, resulting in a value within the range of 1-50 ps^{-1}: this allows to maintain constant the temperature of the simulation while limiting the stochastic effects on the dynamics.

1.2.4.2 Nosé-Hoover

A common alternative to the stochastic approach described above relies on the extended-system method first introduced by Nose [27] and subsequently refined by Hoover [28]. Instead of modelling collisions with a set of fictitious particles, additional degrees of freedom are added to the original Hamiltonian as to represent the heat bath. The extended Hamiltonian adds two additional terms to the equation of motion, corresponding to a thermal reservoir and a friction term: the friction force is proportional to the product of the velocity of each particle and a friction parameter, ξ. The friction parameter is a fully dynamic quantity with its own momentum, p_ξ: the time derivative is calculated from the difference between the current kinetic energy and the reference temperature. The equation of motion of the system are, thus, modified as follows:

$$\dot{\mathbf{p}}_i = \frac{\mathbf{f}_i}{m_i} - \frac{p_\xi}{Q}\mathbf{p}_i, \tag{1.23}$$

$$\dot{p}_\xi = T - T_0, \tag{1.24}$$

where T_0 is the reference temperature of the heat bath, while T is the current instantaneous temperature of the system. The Hamiltonian of the extended system reads as

$$H_{NH}(\mathbf{r}, \mathbf{p}, \xi, p_\xi) = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{r}) + \frac{p_\xi^2}{2Q} + NkT\xi = H_0(\mathbf{r}, \mathbf{p}) + \frac{p_\xi^2}{2Q} + NkT\xi, \tag{1.25}$$

where H_0 is the Hamiltonian of the physical system. Since a conserved quantity exists, it can be shown[3] that the projection of the partition function of the extended system into the original one, leads to recover the canonical probability distribution. In other words, by integrating the equation of motion of the conservative system H_{NH}, for the coordinated $\mathbf{r}$, $\mathbf{p}$, ξ and p_ξ, the NVT ensemble is recovered for the real system of positions and momenta $\mathbf{r}$ and $\mathbf{p}$. A shortcoming of this method is that the Nose-Hoover thermostat can behave non-ergodically for particular systems, since the lack of any stochastic component means that only a subsection of phase space is ever sampled. Moreover, the fictitious mass Q introduced by the thermostat depends on the reference temperature and needs to be chosen carefully in order to avoid the dumping some relevant vibrational frequencies of the system. This issue is normally tackled by replacing the mass parameter with a function of a period τ_T: this is the characteristic period of the oscillations of the kinetic energy between the system and the bath, and has the advantage

[3] Integrating out the fictitious variable of the heat bath, what is left it the canonical partition function

$$Z = \int \xi^{3\mathrm{N}} \delta(H_{NH}(\mathbf{r}, \mathbf{p}, \xi, p_\xi) - E) d\mathbf{r}^N d\mathbf{p}^N \, d\xi dp_\xi = C \int e^{\frac{-H_o(\mathbf{r},\mathbf{p})}{kT}} d\mathbf{r} d\mathbf{p}$$

of being independent from the temperature[4]. The non-ergodic behaviour of a single Nose-Hoover thermostat is solved by employing a chain of consecutive Nose-Hoover thermostats, each one acting on the extended system defined by the previous one; a series of additional friction parameters ξ_i are added with the corresponding coupled equations for the momenta, resulting in a set of coupled equations

$$\begin{aligned}
\dot{\mathbf{p}}_i &= \frac{\mathbf{f}_i}{m_i} - \frac{p_{\xi_1}}{Q_1}\mathbf{p_i}, \\
\dot{p}_{\xi_1} &= (T - T_0) - p_{\xi_1}\frac{p_{\xi_2}}{Q_2}, \\
\dot{p}_{\xi_{i=2,\dots,N}} &= \left(\frac{p^2_{\xi_{i-1}}}{Q_{i-1}} - kT\right) - p_{\xi_i}\frac{p_{\xi_{i+1}}}{Q_{i+1}}, \\
\dot{p}_{\xi_N} &= \left(\frac{p^2_{\xi_{N-1}}}{Q_{N-1}} - kT\right).
\end{aligned}$$

where N is the number of chained thermostats. In the limit of an infinite number of thermostats, the dynamics are guaranteed to be ergodic; a value of N between 4 and 10 is normally considered sufficient.

1.2.4.3 Barostats

Keeping the simulation box constant during MD simulations can sometimes be inconvenient. For the simulation of liquids phases, in fact, the change of the simulation box volume is necessary to adjust the density to his equilibrium value; moreover, it is often desirable to control the pressure of the system to study its behaviour under precise thermodynamic conditions, aimed at matching the experimental ones. The necessity of running simulations in the *isothermal-isobaric ensemble* (NPT) motivates the use of barostatting algorithms. The thermostats presented earlier can be adapted to vary the size of the simulation cell in order to equilibrate the pressure of the system towards a reference external one. A simple method proposed by Berendsen [25] involves the coupling to a pressure bath, according to which the pressure of the system P is proportionally rescaled to the reference pressure P_0 of the bath as

$$\left(\frac{dP}{dt}\right)_{bath} = \frac{P_0 - P}{\tau_p}, \tag{1.26}$$

[4] τ_T is related to Q via the following relation: $Q = \frac{\tau_T T_0}{4\pi^2}$

where τ_p is the relaxation time over which the barostat acts. P, for the case of pair additive potentials, can be simply calculated as the sum of the virial and the kinetic energy term as

$$P = \frac{1}{V}\left[\frac{1}{3}\sum_{i}^{N}\left(\sum_{j>i}^{N}\mathbf{f}_{ij}\cdot\mathbf{r}_{ij} + \frac{\mathbf{p}_i^2}{m_i}\right)\right]. \tag{1.27}$$

The pressure is thus adjusted through a scaling parameters that acts on inter-particle distances $\mathbf{r}_{ij}$ and is applied at each time step. This scaling factor, μ, is calculated as

$$\mu = 1 - \frac{\beta\triangle t}{3\tau_p}\left(P_0 - P\right) \tag{1.28}$$

where β is the isothermal compressibility of the system and $\triangle t$ the time step. The Berendsen scheme allows one to keep the average pressure constant, while the fluctuations are regulated by the ratio β/τ_p. β may not be accurately known in advance, but, in most cases, the use of an experimental value is considered sufficient, since the accurate reproduction of pressure fluctuations might be not necessary, apart from a restricted number of cases, such as for phase transitions. For water at 1 atm and 300 K, β is equal to 4.6×10^{-5} bar, while for most liquids sufficiently accurate measurements can be found in the literature. Values of τ_p between $10^2 - 10^3 \times \triangle t$ are routinely used, while, to recover the correct Bulk modulus with more precision, a finer optimisation is always advisable.

Eq. 1.28 represents the case of isotropic scaling of cell positions; however, the Berendsen algorithm can also be generalised to an anisotropic case. A matrix μ is calculated from the stress tensor $\mathbf{P}$ and multiplied by the particle position in order to scale independently the components of $\mathbf{P}$ to their reference values; the case of orthorhombic symmetry is particularly simple, since all the off-diagonal terms of the tensor are null. Semi-isotropic scaling can also be obtained, which is particularly useful for systems having interfaces (i.e. liquid-solid).

It is important to remind that although the Berendsen pressure coupling algorithm yields a simulation with the correct average pressure, it does not allow to recover rigorously the NPT ensemble. In all those situations in which pressure or volume fluctuations are important, particularly for small systems, the adoption of other methods is advisable. Two of the most widely used methods are the Parrinello-Rahman [29] and the MTTK [30] pressure coupling schemes, which regulate the pressure of the system with the introduction of additional sets of equations for the box vectors and for the cell volume. The similarity of these two barostats with extended variable approach described earlier for the Nos-Hoover thermostats justifies their simultaneous implementation for the generation of the NTP ensemble.

1.3 Solvation models and electrostatics

We mentioned already how, in the field of classical MD, the accuracy of the model used to represent the atomic interactions is fundamental for the characterisation of structural properties of a molecular system: a reliable representation of intra-molecular contributions and an accurate treatment of the electrostatic interactions between solutes and of solute molecules with solvent are needed. The first point has already been discussed at the beginning of this chapter, while long range electrostatic methods were presented in section 1.2.3; however, the importance of solvation and the influence it has on the conformation and aggregation of biomolecules requires some more attention to outline some of the techniques used in the following chapters.

Various computational methods have been developed for the interaction of molecules with the solvent, presenting various levels of trade-off between accuracy and computational cost. The most popular methods can be classified in two main categories: explicit solvent methods, which treat the solvent with full atomic detail, and implicit solvent methods for which, instead, the solvent is replaced by an average potential.

1.3.1 Explicit Solvation

Explicit solvation is by far the simpler and the most widely used solvation method, since it enables a detailed representation of the interactions with solvent molecules. As suggested by the name, solvent and ions are, in fact, represented at the atomic level, allowing for their degrees of freedom to be taken into account in full. Solute molecules and solvent are, therefore, treated in the same way through the combination of *ff*s for solutes and solvents characterised by an analogous potential energy as in eq. 1.4. While various *ff* have being developed for studying fluid phase equilibria and properties of different liquids (e.g. *traPPE* potentials), water is, for obvious reasons, the most common solvent used in MD simulations and several different models of variable complexity and computational efficiency have been proposed to represent liquid water [31].

In most water models, molecules interact via a LJ potential between oxygen atoms, while the Coulomb electrostatic interaction is exerted between charges located at variable number of sites, which are intended to map the electron density around the water molecules. Three-site models are among the most frequently adopted in MD, since they combine the reproduction of experimental properties with high computation efficiency. Charges are assigned to each of the three atom sites, while the application of holonomic constraints (e.g. SHAKE algorithm) to bonds and to the angle formed between the two hydrogens, forces the molecules to maintain a rigid struc-

ture, matching that expected for relaxed water molecules. Despite being introduced sometime ago, TIP3P [32] and SPC/E [33] models are still at present two of the most common three-site models, since they allow to recover with good precision several thermodynamic and structural properties of the liquid phase. The simplified parametrisation of three-sites models, however, presents some limitations at particular thermodynamic conditions (i.e. low temperatures) and for confined water. For this reason, more complex models have been developed, exploiting an augmented charge representation: additional fictitious interaction sites are added to water molecules to account for the lose pair or electrons. TIP4P and TIP5P demonstrated to extend the applicability of water models, providing a correct description of the phase diagram [34], as well as the possibility of studying complex critical phenomena, such as ice nucleation [35].

Explicit account of solvent-solute and solvent-solvent interactions is a desirable characteristic since it leads to a more complete and reliable exploration of the configurational space of biomolecules. It was, for example, demonstrated how the formation of solvation shells and salt bridges, have a profound influence in protein folding [36], thus evidencing the importance of explicit solvation. However, the main drawback is that such high resolution is balanced by a significant computational cost: for typical systems, such as proteins in solution, the number of solvent molecules that is required might add up to more than 60-70% of the total number of particles in the system. This means that explicit solvation significantly increase the computational cost of MD simulations by at least a factor of two (for linear scaling codes). This issue was particularly important in the past for the limitation of computational resources: as a result, continuous solvation models have been developed. In recent years, the high level of parallelisation of modern codes (e.g. exploiting multithreading on GPU cards) makes explicit solvation convenient with respect to continuous models. In addition, coarse-grained water models have been developed to increase performances while providing a comparably sound description of the physical system [37].

1.3.2 Implicit Solvation

In implicit solvent models, the solute molecules are represented atomistically while the solvent is modelled by a continuous medium through a mean field potential. Solute–solvent interactions are, in fact, described by solvation energies ($\triangle G_{sol}$), corresponding to the free energy of transferring the solute from vacuum to the solvent environment. $\triangle G_{sol}$ is traditionally decomposed into three components:

$$\triangle G_{solv} = \triangle G_{cav} + \triangle G_{vdW} + \triangle G_{ele}, \tag{1.29}$$

$\triangle G_{cav}$ is a volume dependent term that accounts for the creation of a cavity within the solvent to accommodate the solute molecule, $\triangle G_{vdW}$ is the non-polar interaction energy of the embedded solute molecule with the solvent and $\triangle G_{ele}$ is the electrostatic interaction term. One popular approximation for the first two terms of eq. 1.29 assumes a linear dependence between the non-polar solvation energy ($\triangle G_{np}^{SASA} = \triangle G_{cav} + \triangle G_{vdW}$) and solvent-accessible surface area ($SASA$) [38]. The free energy of a solute molecule is described by a mean solvation potential computed as the product of an atom-specific solvation energy per surface area, σ_i, and the atomic surface accessible area, $SASA_i(r_i)$, summed over all atoms i as follows

$$\triangle G_{np} = \sum \sigma_i SASA(r_i). \tag{1.30}$$

The parameters σ_i per each atom types are extrapolated by force matching from simulation in explicit solvent [39], while surface accessible area for each particle needs to take into account the overlap with neighbouring atoms [40]. The most common method for the calculation of $SASA_i$ (see [41] for an overview) approximates a solvent molecules to a sphere of radius R_{solv} (typical values for water are 1.3-1.4 Å); this results in a solvent accessible surface of area $4\pi(R_i + R_{solv})^2$, where R_i is the vdW radius of atom i. An alternative approach used for simulation of proteins and nucleic acids, replaces the sum in eq. 1.30, with a linear relation with the total SASA of the solute, via the introduction of a unique surface tension term. SASA method for non-polar solvation have enjoyed surprising success and, despite having several inconvenience caused by the choice of the surface tension parameter and a variable precision of energy estimation, still remains among the most popular technique within available MD codes.

The electrostatic energy term $\triangle G_{ele}$ is determined through the use of a continuum electrostatic model which accounts for the interactions of the solvent with solute's charges through the solution of the Poisson–Boltzmann equations. The electrostatic potential $\phi(x)$ generated by a charge distribution $\rho(x)$ in an environment having dielectric permittivity $\varepsilon(x)$, in fact, is given by the Poisson equation,

$$-\nabla\varepsilon(x)\nabla\phi(x) = \rho(x). \tag{1.31}$$

$\varepsilon(x)$ is equal to the product of the the vacuum permittivity and the local dielectric constant of the system[5]. In order to solve eq. 1.31 the charge density $\rho(x)$ is approximated with the superposition of the contribution of fixed solute charges $\rho_f(x)$ and that of mobile ions $\rho_m(x)$ belonging to the solvent. The solute charge distribution is calculated from the N point charges

[5] Different relative dielectric constants are used for the interior and the exterior of the solute molecule: in the first case common values are between 2-20, while 80 is used for water at room temperature, in solvent-accessible region

Q_i located at each atomic position,

$$\rho_f(x) = \sum_{i=1}^{N} Q_i \delta(x - x_i), \tag{1.32}$$

where δ is a delta function. The mobile charges are instead represented by a continuous "charge cloud" described, at the equilibrium, by a Boltzmann distribution. For M ions of charges q_j having concentrations c_j, the mobile ion charge distribution is given at temperature T by the following

$$\rho_m(x) = \sum_{j=1}^{M} c_j q_j \exp(-q_j \phi(x)/kT). \tag{1.33}$$

Combining these expressions for the mobile ions' charge densities with 1.31, the full PB equation is obtained.

$$-\nabla\varepsilon(x)\nabla\phi(x) = \rho_f(x) + \sum_{j=1}^{M} c_j q_j \exp(-q_j \phi(x)/kT). \tag{1.34}$$

The exponential term $\exp(-q_j\phi(x)/kT)$ can be substituted by the first order Taylor series expansion for $|-q_j\phi(x)/kT| \ll 1$, reducing to the linearised PB equation:

$$-\nabla\varepsilon(x)\nabla\phi(x) = \rho_f(x) + \sum_{j=1}^{M} c_j \frac{q_j^2 \phi(x)}{kT}. \tag{1.35}$$

The solution of eq. 1.35 is obtained via numerical techniques based on the discretisation of the domain applying finite difference or finite elements schemes. Once the PB equation is solved, the electrostatic potential is known for the entire system and the electrostatic free energy can be evaluated by a variety of integral formulations; the simplest, for the linearised PB equation, is

$$\triangle G_{ele} = \frac{1}{2} \sum_{i=1}^{N} q_i \phi(r_i). \tag{1.36}$$

The solution of the PB equation for a molecular system of large dimension and over a set of configurations extracted from the dynamics can be a computationally demanding task. In addition to the PB methods, simpler solvation models have also been derived to deal with continuum electrostatics. Among these simplified methods, *Generalised-Born* (*GB*) is one of the most popular. Sharing the same continuum representation with *PB* method, the *GB* model is based on the analytical solvation energy obtained from the solution of the Poisson equation for a point charge placed in the centre of an ideally spherical solute with effective radius f_{ij}^{GB}

[42]. The molecular electrostatic solvation energy is calculated as

$$\triangle G_{solv} \cong -\frac{1}{2}\left(\frac{1}{\varepsilon_{int}} - \frac{1}{\varepsilon_{sol}}\right)\sum_{i,j}\frac{q_i q_j}{\mathbf{f}_{ij}^{GB}}, \tag{1.37}$$

where ε_{int} internal dielectric for the solute, and ε_{sol} is the dielectric of the continuum solvent. The radius f_{ij}^{GB} for the i,j pair is calculated from the atomic effective radii (R_i) through the originally proposed relation

$$f_{ij}^{GB} = \left[r_{ij}^2 + R_i R_j \exp\left(\frac{-r_{ij}^2}{4R_i R_j}\right)\right], \tag{1.38}$$

where r_{ij} is the pair distance. In practice, the *GB* model uses the individual Born radii R_i for each atomic type by the matching of available solvation energies data for a collection of residues: the availability of such parameters is thus essential for the *GB* method, which under this condition demonstrated to match the results of more detailed models, such as *PB [43]*. In the absence of the correct radii for every molecular species, GB methods fail to match the accuracy of more complete models. Nonetheless, GB methods have become increasingly popular because of their computational efficiency.

The derivatives with respect to the atomic positions of the non-polar and electrostatic contributions describe above, can be inserted into the equation of motion to correct to the force determined for the unsolvated system to account for a implicit medium, and therefore, to run the dynamics coupled with an implicit solvent. These solvation models allowed for the development of the *PB/SASA* or *GB/SASA* methods for the calculation of the biding free energy differences resulting from the interaction between molecules.

The binding free energy of a complex formed by molecules A and B is equal to the difference between the free energy of the complex (AB) and that one of the two species taken separately as

$$\Delta G_{AB} = G_{AB} - G_A - G_B. \tag{1.39}$$

Each individual term of eq. 1.39, corresponds to the sum of the following three terms:

$$G = E_{gas} + G_{solv} - TS. \tag{1.40}$$

E_{gas} is the gas phase energy of the molecule, G_{solv} is the solvation energy and TS is the vibrational entropy contribution. Through eq. 1.39 and 1.40, the binding free energy differences can be averaged over a collection of configurations, sampled from the equilibrated trajectory of a system in explicit solvent. According to the *MM-PB(GB)SA* protocol [44], E_{gas} for the

complex and each individual ligand is calculated upon minimisation of the potential energy; G_{solv} is obtained with one continuum solvation model described above; the vibrational entropy is instead determined by a normal mode analysis upon diagonalisation of the Hessian matrix or, alternatively, with a Principal Component Analysis. Despite suffering the implicit limitation of a continuum solvent model, this methods provide an agile technique for the calculation of free energies that demonstrated to provide consistent results when compared with experimental measurements and other more refined free energy techniques [45].

1.4 Density functional theory

The exact solution of the time independent Schrödinger equation for a many-body system of interacting electrons as in 1.1, is a complex task even for trivial cases. Various theories have been developed to tackle this issue, which are based on a set of approximations that allow one to calculate the electronic structure of an atomistic system with a sufficient level of accuracy and an affordable computational cost. Among these methods, *density functional theory* (DFT) provided a ground breaking tool in the field of materials science and computational chemistry.

DFT, developed by Hohenberg and Kohn in the mid 1960s, introduced a way of mapping a fully interacting many-electron system into an equivalent set of non-interacting single-electron equations moving in an effective potential. The basis for this approximation lies in the fact that the electron charge density is a central quantity in the solution of the eigenvalue electronic problem: the total energy of the system, in fact, is uniquely determined by electronic density $n(\mathbf{r})$; the minimisation of energy with respect to the electron density yields univocally the ground state. The energy of a set of N electrons in an external potential $V_{ext}(\mathbf{r})$ of the fixed ions, can thus be written as a functional of the density:

$$E[n(\mathbf{r})] = F_{HK}[n(\mathbf{r})] + \int V_{ext}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r}. \tag{1.41}$$

F_{HK} is the Hohnberg-Kohn functional, containing the kinetic and electron-electron interaction energies. Exploiting the variational principle with the additional normalisation constraints on the density, the solution of the Schrödinger equation can obtained by minimising the energy functional $E[n(\mathbf{r})]$ with respect to $n(\mathbf{r})$[6],

$$E_0[n(\mathbf{r})] = \min_{n(\mathbf{r})} \{E[n(\mathbf{r})]\}, \tag{1.42}$$

[6]In practice, the minimisation is performed over the densities whose integral correspond to the total number of electrons; this normalisation condition is applied through Lagrange multipliers.

where the E_0 is the ground state energy. However, the analytic form of $F_{HK}[n(\mathbf{r})]$ is not known for a system of interacting electrons. Khon and Sham proposed to solve this by replacing the interacting problem with a fictitious system of N non-interacting orbitals $|\phi_i\rangle$ yielding the same density, as

$$n(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2 . \tag{1.43}$$

According to this single particle scheme, $F_{HK}[n(\mathbf{r})]$ can be expressed as the sum of kinetic energy of non-interacting electrons, which can now be written as

$$T_s[n(\mathbf{r})] \;=\; -\frac{1}{2}\sum_i \int \phi_i^*(\mathbf{r})\nabla^2\phi(\mathbf{r})d^3\mathbf{r}, \tag{1.44}$$

while the Hartree electrostatic contribution arising from the electron charge density is equal to

$$E_H[n(\mathbf{r})] \;=\; \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d^{\mathbf{3}}\mathbf{r}d^3\mathbf{r}'. \tag{1.45}$$

An additional term, $E_{XC}[n(\mathbf{r})]$, is necessary to be added to account for the missing exchange and correlation energies. The minimum condition of eq. 1.42 reads now as

$$E_0[n(\mathbf{r})] = \min_{n(\mathbf{r})} \left\{T_s[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{ext}[n(\mathbf{r})] + E_{XC}[n(\mathbf{r})]\right\}, \tag{1.46}$$

where the last three term represent the Kohn-Sham functional. Writing the functional explicitly in terms of the density built from non-interacting orbitals $|\phi_i\rangle$ (eq. 1.43) and applying the variational theorem, the original eigenvalue problem results in a set of N single particle equations,

$$\left(-\frac{1}{2}\nabla^2 + \underbrace{V_H(\mathbf{r}) + V_{ext}(\mathbf{r}) + V_{XC}(\mathbf{r})}_{=V_{KS}(r)}\right)\phi_i(\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r}) \qquad i = 1, ..., N, \tag{1.47}$$

where V_H and V_{XC} are derived by taking functional derivatives of E_H and E_{XC}, and V_{KS} is the Kohn-Sham effective potential. These non-linear equations are solved self-consistently via iterative methods and allow one to determine the ground state density and the ground state energy of the system. In principle, the eq. 1.47 are exact; however, a reliable form for the exchange and correlation potential V_{XC} is needed. This functional is, in fact, the only uncontrolled approximation not explicitly treated in the DFT formalism, containing all the many-body effects; the development of approximations to the functional has been an active topic of research since DFT was first proposed.

Two of the most common approaches are the *Local Density Approximation* (*LDA*) and *Gen-*

eralised Gradient Approximation (*GGA*). The first method assumes that E_{XC} is equal to that of a homogeneous electron gas and depends only on the local charge density: LDA functionals originally proposed for simple metals and semiconductors, find application in a wide range of system in condensed matter, while the systematically overestimation of the exchange energy leads to poor results for molecules. GGA functionals extended LDA ones by including the dependence on the local gradient of the charge density, resulting in a more robust approach for systems where the electron density varies rapidly. GGA functionals guarantee a better representation of binding energies, at the expense of a moderate computational cost; moreover, they provide an improved description of strong inter-molecular interactions such as hydrogen bonds, whose effect are significantly overestimated by LDA. Some implementations of GGA functionals combine different correction methods for the exchange and correlation part in order to match more closely the experimental data for a broad range of atomic and molecular systems; these are referred to as hybrid functionals, with the BLYP functional being the most widely used example.

Calculating the interaction of molecular species distant just a few Angstroms requires dispersion forces to be taken into account. Local (LDA) and semi-local (GGA) density functionals fail to represent vdW interactions: their intrinsic mean field approach, in fact, cuts the long tail contribution of the charge density. For this reason LDA and GGA are consider to yield the upper and the lower bound for the energies involved in the formation of complexes of molecules and their physisorption onto a metal surface. In recent years, van der Waals functionals have been developed to include explicitly non-local contributions of the correlation energy in terms of the formulation of functionals which include density-density interactions: the DFT-DF scheme [46] allowed to calculate the dispersion forces between weak interacting noble atoms and provided the starting point for the development of further improvements, based on self-consistent corrections [47].

1.4.1 The basis set

In order to solve eq. 1.47, the wavefunctions need to be represented through a convenient basis set. The choice falls into two categories: atom centred basis sets, such as gaussians, or plane waves, which offer the same accuracy throughout the system. Localised basis sets are preferred for molecular systems since the implicit non-periodic nature of the charge density implies that far fewer functions are needed when compared to plane waves to reach an analogous precision; for this reason, localised basis sets are popular in quantum chemistry since they facilitate the calculation of the interaction between biomolecules. For periodic system, instead, plane waves

are a more convenient choice as they provide a more straightforward representation. The reason for this lies in the fact that for a condensed system, the electron wavefunction are subjected to the Bloch theorem, which holds for particles in a periodic potential $V_{ext}(\mathbf{r}+\mathbf{R}) = V_{ext}(\mathbf{r})$, where $\mathbf{R}$ is a vector of the Bravais lattice. According to this theorem, the wavefunction can be split into the product of an exponential term and periodic function, having periodicity equal to the periodicity of the lattice. The latter can be expanded as a Fourier sum over the reciprocal lattice vectors G of the periodic system,

$$\phi_{i,\mathbf{k}} = \sum_{G} c_{i,(\mathbf{k}+\mathbf{G})} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{1.48}$$

where the $c_{i,\mathbf{k}}$ are the expansion coefficients for each band and for each $\mathbf{k}$. The sum is performed over a set of reciprocal vectors for which $\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 < E_{cut}$: this corresponds to the introduction of an kinetic energy cut off, which put a limit to the number of plane waves used to approximate the wavefunction, provided that the ground state energy is correctly calculated, within DFT precision. In principle, to determine the physical properties of a periodic system, it is necessary to integrate over all $\mathbf{k}$ in the first Brillouin zone. However, it is sufficient to use a finite k-point sampling scheme, since for aforementioned periodicity the electron density is a smoothly varying function of $\mathbf{k}$: as a consequence a finite number of k-points is selected from a uniform discretisation of k-points, allowing to approximate the band structure (e.g. Monkhorst-Pack grid [48]). For a large system, the volume of the reciprocal unit cell is very small, thus a smaller number of k-points is necessary to adequately sample the Brillouin zone. In the limit of a very large system, such as for isolated molecules in the gas phase, we can consider only a single point, for which $\mathbf{k} = \Gamma$; a sufficiently large cell with enough vacuum-padding allows to separate periodic images, thus preventing spurious interactions between them.

1.4.2 Pseudopotentials

The computational cost of DFT is related to the number of plane waves used to represent the wavefunctions, and thus depends on the value of kinetic energy cut-off (and on the volume). However, close to the nucleus, the space confinement of core electrons causes the wavefunction to oscillate widely in agreement with the orthogonality condition. As a result, a very large number of plane-waves would be required to describe these oscillations with the same precision as elsewhere. This issue was tackled by replacing the true potential exerted by core electrons and ions with a weaker, smoother pseudo-potential. The result of this operation is that the pseudo-wavefunction calculated as such can be conveniently represented by a smaller number of plane waves, so that the computational cost is significantly reduced.

The use of pseudo-potential is supported also by some physical considerations: the core electrons, in fact, do not always need to be represented explicitly since they are expected to marginally influence the chemical properties of the system under investigation; correspondently, the wave functions for the valence electrons are approximated only within a limited region of space close to the nuclei, while elsewhere they have a much smoother shape. Clearly, this approximation holds at the condition of yielding the same valence spectrum with respect to the all-electron case; additionally, pseudo-potentials need to be transferable.

Two are the most common types of pseudo-potential used in DFT. On one side, it is common to require that the core states of the all-electron system and pseudo-wavefunctions within the core radius contain the same number of electrons; from this condition are derived *norm-conserving* pseudo-potentials. The relaxation of the norm-conservation constraints allow to formalise an alternative class of pseudo-potentials, developed by Vanderbilt and known with the name of *ultrasoft* pseudopotentials: they are smoother than norm-conserving pseudo-potentials, and are particularly used since they require even fewer plane waves i.e. a lower cut off energy.

1.4.3 Force evaluation

For a set of nuclear position the solution of the Kohn-Sham equations is carried out self consistently from an initial guess of the charge density: the resulting eigenfunctions are iteratively used to update the density until convergence. Once the electronic state of the system is defined, the forces on each atom can then be calculated thanks to the Hellman-Feynman theorem [49] which simplifies their evaluation. Given that the derivative of Kohn-Sham wavefunctions vanishes, the force on *i-th* nucleus is simply given by the derivative performed over the Kohn-Sham energy operator:

$$\mathbf{f}_i = -\frac{\partial E_{KS}}{\partial \mathbf{R}_i} = -\sum_j \phi_j | \frac{\partial H_{KS}}{\partial \mathbf{R}_i} | \phi_j, \tag{1.49}$$

where f_i is the force acting on the atom of position R_i, and the sum is performed over all bands. Once these forces have been calculated, they can be used for the integration of the Newton's equation of motions. Alternatively, geometry optimisation procedure can be performed, by noting that forces on atoms have to be null for a relaxed configuration. The most used methods for finding the minimum in a multidimensional space are the Conjugate Gradient algorithm and quasi-Newton methods, such as *BFGS*.

1.5 Minimisation algorithms

To perform geometry optimisation on a molecular system described by a classical potential or through DFT, the minimisation of the total energy needs to be performed on a 3N dimensional space of the atomic position **R**. Efficient algorithms are required in order to converge to the global minimum in the smallest number of iterations, independently on the starting point and, possibly, without being trapped in local minima. The most widely used minimisation algorithms differ in complexity and efficiency. The simplest one is *steepest descen*t method: the minimum is approached by simply following downhill the direction defined by the gradient of the potential

$$\mathbf{r}_{i+1} = \mathbf{r}_i - h\nabla V, \tag{1.50}$$

where h is the step size. Despite its simplicity, this algorithm is not frequently used since it does not perform efficiently for non-smooth PES, having significantly different curvatures along different directions. As a result, convergence might not be easily reached within a convenient number of iterations. For this reason, the *Conjugate gradient* algorithm is often preferred: the search for the minima is performed along conjugate directions (through the gradient), in such a way that the solution is more efficiently reached and instabilities are reduced. Typically, the number of step needed for a minimisation is reduced with respect to steepest gradient, unless conditioned strategy are adopted. An more efficient and precise method is the Broyden-Fletcher-Goldfarb-Shanno algorithm (*BFGS*) which is based on a quasi-Newton approach employing an approximated reconstruction of the Hessian matrix to update iteratively the guess at each step.

Also MD can be used to performed energy minimisation in a straightforward way. Fictitious forces proportional to the particles' velocities are applied to quench the systems to an energy minimum. This is done, usually, by performing short MD runs starting from an initial high temperature, which allows to overcome the local energy barriers that might otherwise hinder the reach of a global minimum. The global minimum, indeed, is not always found but an optimal choice of the damping coefficient γ (see eq. 1.21) can greatly improve the convergence. This methods that be easily used in conjunction with a stochastic thermostats and is particularly suited for finding the minimum of system characterised by a complex PES. A simple implementation of this method is obtained by scaling the target temperature of the thermostat in a linear way over the course of few steps of dynamics: this is procedure repeated a number of times for a different set of randomly chosen initial velocities, to compare the different results obtained.

1.6 Softwares

We report here some information about the softwares and other tools that have been used to carry out the simulations and to analyse the results presented in the next three chapters. *Amber 11* [50] and *Lammps* [51] are the main two software distributions used for classical MD simulations. The molecular structures for all system have been created through the *Avogadro* molecular builder [52], while the parametrisation of the molecular models with the selected *ff* have been carried out using the *Leap* tool included in the Amber tool package. *Quantum Espresso* [53] was used for the DFT calculation reported in chapter 2, while *Vasp* 5.3 [54] was preferred for the relaxations of chapter 3.

Thermodynamic outputs and trajectories have been mainly analysed using Python scripts. The Monte Carlo models presented in chapter 2 was also implemented in Python, which was used to call parallel instances to external executables, of both *Lammps* and *AMBER*. All the plots contained in this manuscript have been also created through *Matplotlib*, a Python visualisation module.

Chapter 2

Chiral dependent gelation of poly-aromatic LMOG

In this chapter, we report on the study carried out to characterise the interactions between a particular type of chiral binaphtol derivatives, in the attempt of explaining the observed formation of a stable gel upon sonication of a solution containing an enantiomeric excess. We found that these molecules have two main conformations which bind in a different way when having opposite chiralities, thus resulting in significantly different binding energies. Importantly, our calculation also allowed to identify a dimer structure that we propose to be one of the possible building block leading to fibre formation: this hypothesis is motivated by the different binding energy calculated for heterochiral dimer with respect to homochiral ones; moreover, we showed that homochiral dimers are able to aggregate in a directional and repetitive fashion, while the same possibility was not observed for heterochiral dimers and those formed by a methylated derivative. This directional hypothesis was tested with a simple Monte Carlo lattice model, which evidenced a decreasing stability of aggregates for decreasing enantiomeric excess; a decrease of the critical temperature is also expected, favouring a more disordered state at room temperature.

2.1 The gelation of binaphtol derivatives

Gels are particularly interesting colloid materials formed by molecules able to create a solid-like network within a quasi-liquid phase. The combination of properties of solid materials with those typically exhibited by liquids has motivated the interest that gels generated in the

scientific community for their potential applications. Nowadays, gels are common materials widely used in the pharmaceutical and cosmetics industries. However, intense research in the field of nanotechnology, particularly in the context of self-assembly, has fostered the attention towards gel formed by small compounds known for this reason as *Low Mass Organic Gelators* (LMOG). The possibility of tuning the molecular building blocks to control the organization and the characteristic of supramolecular aggregates such as gels is indeed a promising strategy to create complex structures bearing specific chemical/physical properties at the macroscopic level. A comprehensive review of recent achievements in the application of gels in medicine and technology can be found in [55, 56].

A group of experimental collaborators at the University of Namur, while working on particular type of interesting binaphtols derivatives, observed the possibility of turning a solution of such molecules into a gel through sonication. The formation of the gel was not expected nor intentional: in the attempt of achieving a finer mixing the molecule through the propagation of sonic waves within the liquid, instead of achieving an emulsion, the solution surprisingly transformed in a few seconds into a viscoelastic substance (see fig. 2.1).

This experimental finding is interesting for various reasons. First of all, the poly-aromatic molecules responsible for this case of gelation are unusually small. Normally, gels are formed by molecules of bigger size, like polymers: large molecules easily form a high number of non-bonded interactions that can induce the formation of a matrix of interlaced fibrils able to trap the solvent. For small compounds, this number is decreased considerably: strong bonds can still form with a repetitive feature, leading preferentially to organic crystal. A gel can instead form if the geometry of the molecule is such that directional interactions [57] cause the aggregation of molecules in fibres, further stabilised by the solvent. This is due to formation of multiple non-covalent interactions, typically in the form of hydrogen bonds or $\pi-\pi$ stackings; compounds with a high content of hydrogen bond donors and acceptors, like carbohydrates and peptides, and aromatic groups, such as steroids, are often found to be LMOG [57].

The second reason of interest relies on the the fact that the gel is sensitive to the optical activity of the solution. The binaphtols derivatives responsible for gelation are indeed chiral compounds with two chiral centres. It was observed that the gel formation was achived only for an enantiomeric pure composition: the introduction of an increasing fraction of the opposite enantiomers, by progressively decreasing the enantiomeric excess in the solution, yielded smaller volumes of gel; for a racemic mixture, the gel was not essentially formed. This feature is undoubtedly interesting: while chirality is an important factor in chemistry, it has rarely been observed having such a clear and profound effect on gelation. Only a few studies, in fact, pointed out independently that specific chiral compounds may exhibit different aggregation behaviour

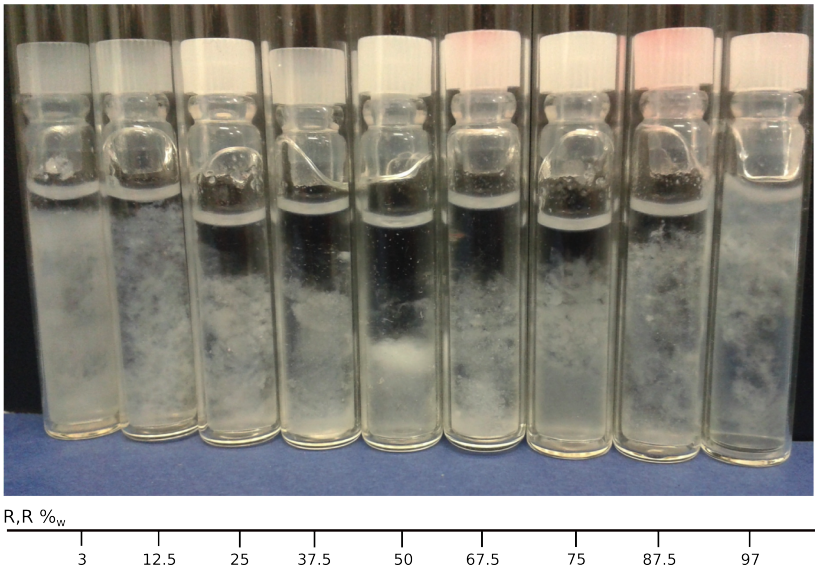


Figure 2.1: Vials containing the gel formed by an increasing fraction ($\%_w$) of the R,R binaphtol derivative in cyclohexane solvent.

[58, 59, 60].

Lastly, the gel was formed only upon sonication. This is quite an unusual fact, since gels normally are obtained by heating and subsequently cooling a solution containing a certain amount of gelators. Also in this case, few accounts of a similar occurrence are found in the literature. Notably, the interest that sonochemistry attracted in recent years in the field of green chemistry [61], enhance the importance of finding further examples that can help to provide a better understanding of the role of sonication for industrial applications.

2.1.1 The binaphtol derivatives

The molecules displaying the interesting gelation property briefly described above belong to a class of poly-aromatic molecules that is synthesised upon the reaction of binaphtol groups with a benzene ring by means of nucleophilic substitution reaction. The binaphtol precursors, designed more precisely with the name *1,1'- bi-2-naphtol*, are molecules frequently used in organic chemistry as ligands for the formation of metal-organic complexes. A distinctive characteristic of this compound is its axial chirality: the bond formed between two binaphtyl groups (see left of fig. 2.2 a) is responsible for the existence of two stable enantiomers, denoted with the notation as R and S to identify the anti-clockwise and the clockwise specific rotation of $+/-$ 35.5^o, respectively. The reaction of the binaphtol groups with the benzene rings leads to various molecular structures depending on the specific reaction mechanism and on the type of precursors

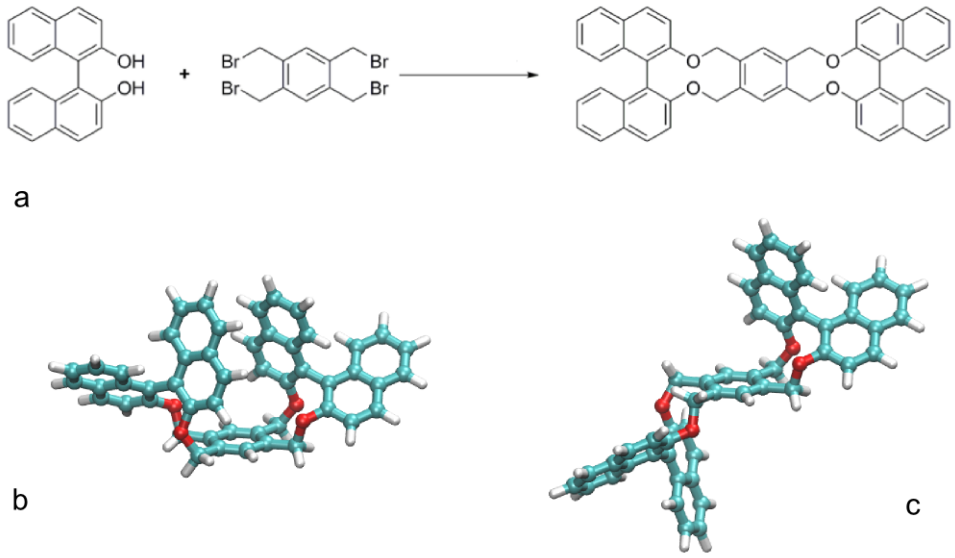


Figure 2.2: Reaction mechanism of the binaphtol derivatives from the precursos *(1,1'- bi-2-naphtol)* (a). The molecules were found having two conformations, the folded and the extended ones, labelled with the letters b and c, respectively.

used. Up to three binaphtol groups can be linked to the benzene, but the resulting molecule is not considered here for it was not found to induce gelation. On the contrary, this phenomenon was verified experimentally when the binaphtol substituents are symmetrically placed on the opposite edges of the aromatic central moiety, as depicted by the reaction scheme in fig. 2.2.a. Given that each precursor can be either the R or the S isomer, a total number of three possible stereoisomers can be formed: the RR and SS racemates, in the case of symmetric reaction, or the RS (an the mirroring SR) atropoisomer, from an asymmetric synthesis. Despite having the same chemical composition and a similar molecular structure, the atropoisomers tended to perform differently as far as the gelation mechanism was concerned. Dispersion of the RS and SR isomers, indeed, did not exhibit any evidence of gel formation even after sonication: a suspension was formed instead, followed by precipitation. On the contrary, the homochiral isomers RR and SS, did form a gel if used separately, but not when both are present in the dispersion at the same time, yielding the aforementioned sensitivity to the chiral composition. Although the ineffective gelation for the RS and SR isomers could provide some hints regarding the mechanism of aggregation, this work will be focused on the enantiomers RR and SS only.

Given the symmetry of the binaphtol molecule, the two isomers RR and SS can assume two main configurations depending on the relative positioning of the binaphtol groups with respect to the plane containing the central benzene ring. The naphtol substituents can, indeed, be found on the same side, in a *cis*-like isomer, which will be called *folded* henceforth, or on the

Solvent	R,R [mg]	Result	Dipole moment [D]
Cyclohexane	0.9	Transparent gel	0
Pentane	1.1	Precipitation	0
Benzene	1.3	Solution	0
n-Hexane	0.9	Solution	0.08
Toluene	1.2	Precipitation	0.45
Dicloro-methane	1.0	Solution	1.14
Chloroform	1.0	Solution	1.15
Diethyl-ether	1.0	Precipitation	1.15
Ethanol	1.0	Weak gel	1.66
Methanol	1.1	Opaque gel	1.7
Acetonitrile	1.1	Precipitation	3.44
Nitromethane	1.2	Precipitation	3.56

Table 2.1: Tested solvent for the gelation of binaphtols.

opposite side of the benzene group, for the *trans*-like isomer named *extended* (see fig. figure 2.2 *b* and *c*, for the folded and the extended configurations, respectively). It was not clear whether the existence of the two configurations is fundamental for the gel formation. Some preliminary NMR studies were not, in fact, able to provide any insight into the possible role played by each of the two isomers for the formation of aggregates. The results were only able to confirm that both conformations were found in solution before and after gelation but it was not possible to discern the composition of the gel in these terms.

A picture of the system is not complete without considering the solvent. Even though the binaphtol molecules are undoubtedly responsible for an aggregation pattern leading to the formation of a network providing the gel with a finite resistance from shear stresses, the main component of a gel in ponderal terms is nonetheless the solvent. This change in the macroscopic properties of the solution is indeed achieved upon gelation just for a small fraction of gelator molecules: 0.22 $\%_w$ was the minimum quantity of gelator observed to induce the formation of a gel in cyclohexane (*Gelation Critical Concentration*). Being the major component of the gel, the solvent is thus an important factor to be taken into account for understanding the gelation mechanism [62]. The interactions of the solvent with the gelator are fundamental because they influence the gel formation in various ways. Notably, the solubility of a gelator in a given solvent has to be considered because it determines the mixing level of the solution, which is necessary to initiate the growth of the aggregates. If gelators are not soluble in the solvent, for example, they can separate from the solution as precipitates. Conversely, if gelators are highly soluble, an increased affinity with the solvent molecules might stabilise gelators in solution, hindering their aggregation. As a consequence, the stability of the aggregates constituting the gel matrix

is also determined by solvation energy contributions.

The reported gelation phenomenon was attempted for various compositions of the solution. The gelation was achieved for just three solvents, cyclohexane, methanol and ethanol; in all the other cases, sonication resulted in either an unchanged solution, or the precipitation of the solute. The composition of the various system studied experimentally is shown in table 2.1.

2.2 The experimental findings

2.2.1 Analysis of NMR measurements

NMR is a widely used experimental technique for carrying out measurements based on the spin relaxation of nuclei that can be directly related to the microstructure of a material. NMR is, in fact, frequently used to evaluate the size of cavities present in porous materials, where the inclusions are filled with liquids having different magnetic properties. This investigation technique relies on the fact that magnetic decay of the substance contained in a single pore can be described by an exponential:

$$M(t) = M(0)e^{-t/T_2}, \tag{2.1}$$

T_2 is the spin relaxation time and is function of the properties of both the fluid filling the pore volumes and the wall's material. T_2 can be related to the previous quantities by the following relation

$$\frac{1}{T_2} = \frac{1}{T_{2b}} + \rho\frac{S}{V}, \tag{2.2}$$

where T_{2b} is the relaxation time of the bulk fluid, ρ is the surface relaxation strength, and S/V is the surface-to-volume ration of pores. In the limit of restricted diffusion, when, for instance, the pores are not interconnected, each decay process can be considered independent from other events occurring in neighbouring pores. The overall magnetic decay results from the sum of exponential terms, essentially dependent on the surface-to-volume ratio, i.e. to the radius of the pore. If the distribution of the pore size is highly homogeneous, then the measure relaxation times are proportional to the average pore radius (the geometric mean if often considered)

$$M(t) = M_0 \sum_i a_i e^{-t/T_{2i}}. \tag{2.3}$$

An analogous technique was used to estimate NMR relaxation times from gel's samples of varying compositions. The two chiral atropoisomers were mixed at different concentration, varying the composition of the solution from $0\%_w$ to $100\%_w$ of the RR enantiomer fraction. After sonication,

R,R fraction	Mean measured T_2^m $[ms]$
0.000	1611
0.060	1726
0.125	1696
0.250	1735
0.357	1749
0.500	1834
0.625	1770
0.750	1717
0.875	1680
0.940	1694
1.000	1611

Table 2.2: Mean T2 relaxation times measured after the gel maturation for an increasing fraction of the R,R enantiomers. The error on the measurements correspond to about 19 ms.

samples of resulting microstructure (gel if formed) were analysed by NMR: the measurements were repeated until the relaxation times stabilised to their final values (after about 10 days). The results are reported in table 2.2.

Through eq. 2.2, the measured relaxation times provide information on the structure of the gel: the higher T_2^m, the bigger is the typical size of the liquid inclusions. The two limiting values correspond to the relaxation time of the solution (i.e. prior to sonication) and the one for an enantiomeric pure gel. The latter corresponds to the first and last lines of table 2.2, while, for the solution, it was found to be equal to 1838 ms. The data of table 2.2 indicate that a decrease of the enantiomeric excess corresponds to an increased relaxation time. From $T_2^{Gel} = 1611$ ms reported for the homochiral gel, T_2^{Sol} reaches a values of 1833 ms, indicating that the system did not change much upon sonication (cfr. with 1838 ms found for the solution), thus no gel is formed. These data can be used to estimate the quantity of gel present at a given enantiomeric fraction, assuming that the measured relaxation times are a combination of the relaxation times of the gel and the solvent weighted by their solvent fraction X_S[1].

In fig. 2.3 the solvent fraction X_S is reported as a function of the R,R fraction. The first thing to notice is that the plot is characterised by an evident symmetry; this fact is fairly easily explained by the fact that, since the two chiralities are equivalent, it is reasonable to assume that they exhibit analogous results for symmetric compositions. For an enantiomeric

[1]The measured relaxation time can be expressed as $T_2^m = XT_2^{Sol} + (1 - X_s)T_2^{Gel}$, where the solvent fraction can be derived as

$$X_s = \frac{T_2^m - T_2^{Gel}}{T_2^{Sol} - T_2^{Gel}}$$

pure composition, corresponding to $0\%_w$ and $100\%_w$ of R,R fraction, no free solvent is reported, meaning that the totality of the solution was turned into a gel. As the R,R fraction increases (or equivalently decreases from $100\%_w$), the residual solvent fraction increases, as a result of deficient gelation. Importantly, the increment of the solvent fraction is rather marked and just for $6\%_w$ of enantiomeric fraction, the quantity of solvent reached almost a value of $40\%_w$. For other compositions, the solvent fraction varies between $30\%_w$ and $50\%_w$, while the solution containing both atropoisomers in equal concentration resulted in no gel formation. It is worth noticing that for the fraction equal to $12.5\%_w$ and $87.5\%_w$ a small decrease in the solvent was evidenced; this was observed on both sides of the plot suggesting that a different aggregation mechanism might take place for this enantiomeric excess.

The picture emerging from the data of both table 2.2 and fig. 2.3 suggest that the presence in solution of mirroring stereoisomers inhibits the formation of the gel. In particular, there seems not to be a large difference in the effect the atropoisomers have over a broad interval of compositions. Comparatively, the gel formation seems indeed to be extremely sensible to just a small ratio of atropoisomers, as little as a $6\%_w$ fraction.

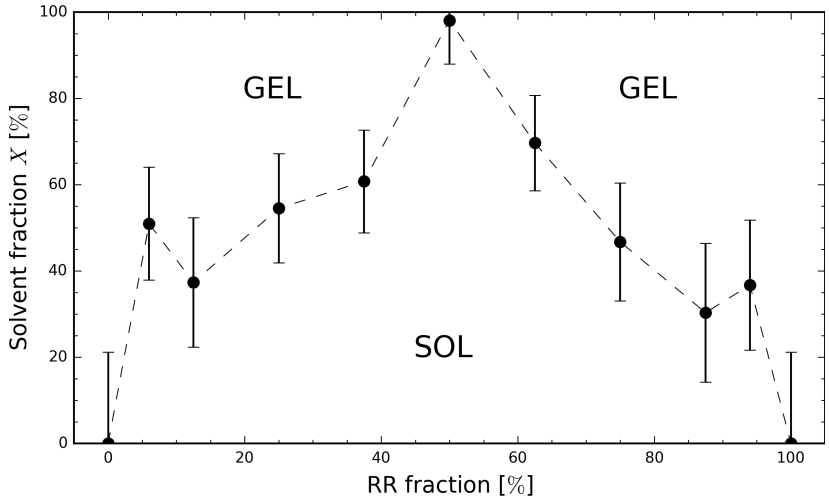


Figure 2.3: Solvent fraction calculated from the relaxation times as a function of the fraction of R,R enantiomers in solution ($\%_w$).

2.2.2 Gel micrographies

It is essential, in order to complement the information obtained by NMR measurements to carefully examine the available micrographies of the gel microstructure. Upon solvent extraction, some samples of the dried gel were examined through SEM microscopy to characterise the shape, the size and the main features of the aggregates. This analysis can give valuable qualitative

information on the physical system that is necessary to create a computational model aimed to provide insight into the intrinsic complexity of supramolecular aggregation.

There are some important considerations to keep in mind about the imaging of soft materials. Firstly, micrographies alone are not enough to infer the molecular basis for the gelation process: the resolution achievable for a disordered, three-dimensional system such as a gel, does not allow to resolve the aggregation structures with atomistic detail. It is also necessary to consider that it is not possible to examine the real structure of the gel as formed: only a sample of the dried gel can be examined after the solvent has been extracted though evaporation. What is left, in fact, is a matrix of interlaced fibres of the gel after drying. Depending on the solvent extraction methods used and the properties of the solvent itself (e.g. viscosity and surface tension), the induced viscous stresses and capillarity pressure can significantly alter the gel morphology, causing deformations of the pre-existent gel matrix. This phenomenon, generally referred to as shrinking, may as well determine the collapse of the gel structure [63].

We mentioned already in a previous section how the binaphtol derivatives are capable of undergoing a sol/gel transition upon sonication with a chiral selective mechanism. Two SEM micrographies of the aggregates found before sonication in a enantiomeric pure solution of gelators in cyclohexane are shown in fig. 2.4. The aggregates appear as spherical vesicles of approximately 500 nm of diameter. Their surface is covered by smaller aggregates one order of magnitude smaller, around 50 nm; it is not clear whether their presence is due to the sample preparation or if they are a characteristic feature of the aggregates. The spherical vesicles, however, are hollow: this can be observed in both the left picture of 2.4 and in the detail shown on the right. An opening in the vesicle, probably caused by the evaporation of the inner solvent, allows one to determine its internal structure. The aggregates are not bulky spheres, rather they are spherical shells, whose 10-20 nm thick membrane incapsulates a certain volume of solvent. The vesicles also reveal an irregular surface texture: irregular bumps of a few nm in size can be distinguished within the membrane, almost as if it was itself formed by smaller aggregates.

A sample of the microstructure of the gel formed upon sonication of a solution of the same composition as before was extracted and imaged; the micrographies reported in fig. 2.5 show a surprising change in the shape of the aggregates. In place of spherical vesicles, sonication induced the formation of fibres. The fibres have a uniform size of about 50-100 nm thickness and a length probably exceeding a few micrometers. At a higher resolution, the surface of each fibre can be examined in more detail. The filaments, clearly visible in the left picture of 2.6, appear as branched structures interlaced with each other to form a complex network. Their surface is irregular and characterised by pores and fracture lines that split the fibre (again, this can be caused by the shrinking induced by solvent extraction before imaging). In the right picture of

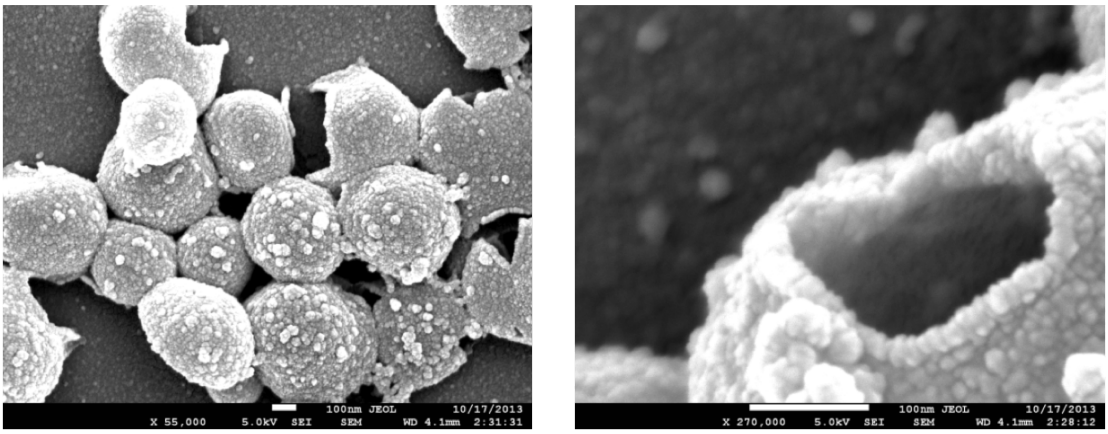


Figure 2.4: SEM micrographies of the spherical aggregates formed by R,R binaphtol derivatives in cyclohexane before sonication (left); a detail of the vesicle shows a shell thickness of about 10-20 *nm*.

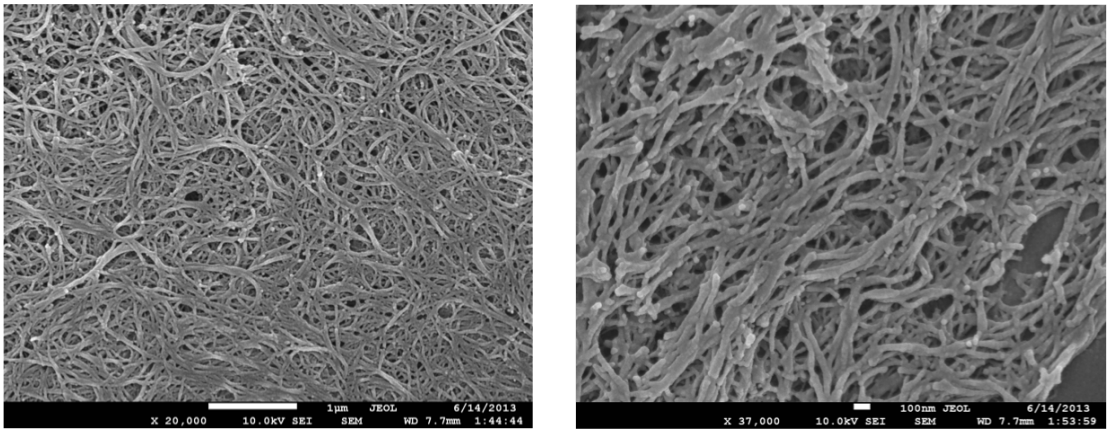


Figure 2.5: SEM micrographies of the fibres formed by sonicating the solution of R,R binaphtol derivatives in cyclohexane.

fig. 2.6 one of the spherical aggregates of fig. 2.4 can be distinguished among the fibres. In particular, this micrography allows one again to notice how the porosity define smaller spherical aggregates of 10-20 *nm* of size, analogously to what was already shown in fig. 2.4. These first micrographies suggest that the distinct form of the aggregate is nonetheless characterised by the aggregation of smaller nanometric sized clusters. Moreover, sonication seems to change the geometrical feature of aggregates from spherical detached vesicles to a matrix of highly interlaced filaments.

In fig. 2.7 are reported two TEM micrographies corresponding to the two types of aggregates presented herein. On the left, one vesicle analogous to those in fig. 2.4 is shown. From these images we can have further evidence of the fact that the particles are indeed hollow (the darker margins of the particle when compared to the inside indicate a distinctively less dense

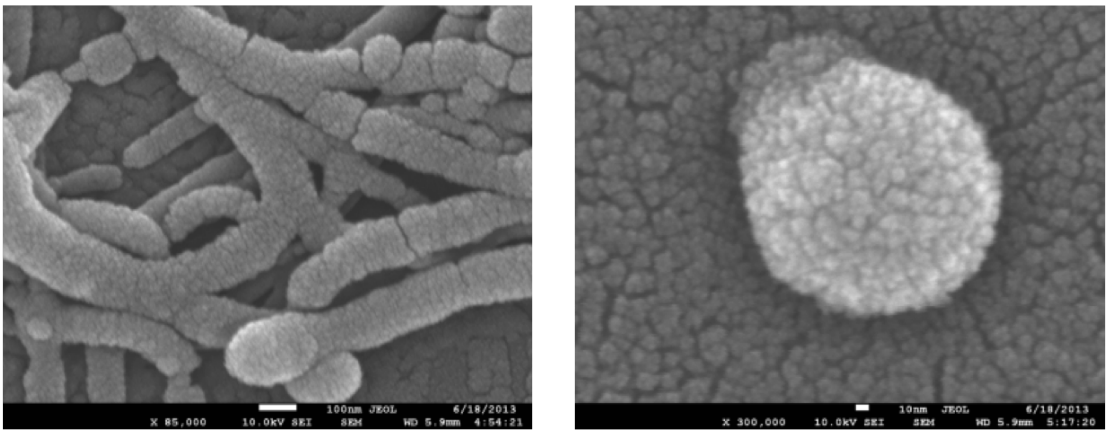


Figure 2.6: SEM micrographies of the fibres: a detail of the interlaced fibres network (left); a survived spherical vesicle appear to be formed by smaller particles.

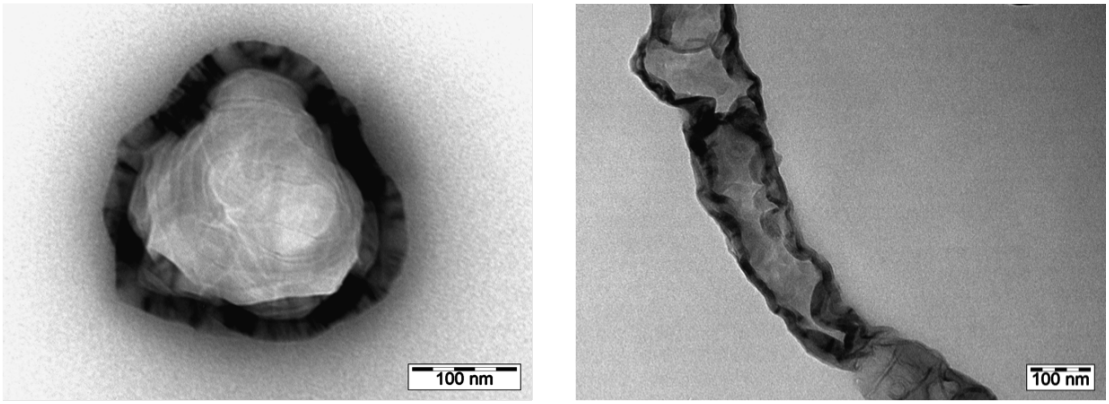


Figure 2.7: TEM micrographies of the aggregates: hollow vesicles (right); hollow fibres formed through sonication (left).

volume), they have a size of few hundreds nanometers and have a wall thickness of 10-20 *nm*. Surprisingly, this vesicle was detected in a system after sonication, meaning that not all the vesicles transform into fibres and can still be present in the gel. A TEM micrography of one of the fibres is shown on the right: also these aggregates are expected to accommodate solvent in their inner part, resembling a sort of tubular duct. The hollow fibres are far from being a straight cylinder, though: a swelling in the top part of the fibre in fig. 2.7 is reminiscent of a spherical aggregate. Similar evidence can be found also in the lower part of left image of fig. 2.6. These two observations, together with the irregular cross section evidenced by the TEM micrography, led us to formulate the hypothesis that the fibres might be formed by the aggregation of vesicles followed by the coalescence of their shells, similarly to the fusion of lipid membranes.

The effect of the chiral sensitivity of the gelation mechanism can be better appreciated by the

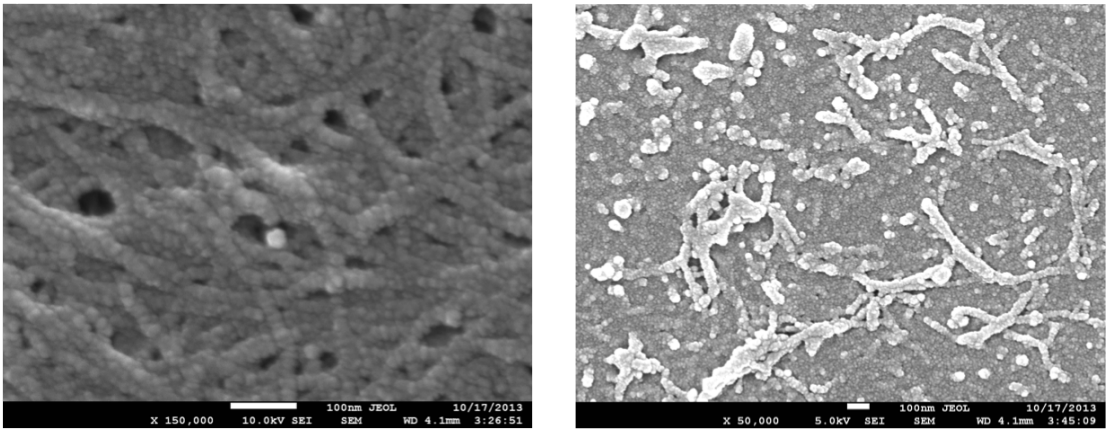


Figure 2.8: SEM micrographies for the heterochiral compositions: the microstructure formed by a solution containing a R,R fraction of 70%$_w$ in cyclohexane still shows some fibres (left); for the racemic solution, just few fibres are found (right).

SEM micrographies of fig. 2.8. The presence of a fraction of enantiomers higher than 20%$_w$ progressively inhibits the fibre formation to the point that, above 35%$_w$, no fibres are formed but molecules remain in solution. On the left of fig. 2.8, in a solution of R,R binaphtol derivatives, a fraction of 30%$_w$ of enantiomers SS was added. The resulting microstructure still keeps a trace of a fibrous network, similar to the one reported for the enantiomeric pure case, but it is less defined. This is consistent with the higher solvent fraction reported by NMR measurements. For a racemic solution, when both chiralities are present in equal concentration, the fibres almost disappear. Again, this is consistent with the NMR measurements that for this composition do not report any gel formation.

Among all the series of gelation experiments conducted and images taken of the resulting nanostructures, only the most relevant were considered here. Intermediate cases were left out on purpose given the complexity of the problem. This is the case, for example, in systems where both the two chiralities are present within the same molecule at the same time (RS-binaphtol). Additionally, vesicles and fibres are found in some measure in nearly all the systems considered, while here only the most representative have been selected and presented in this section.

2.3 Ultrasounds in chemistry

The activation process that was found to trigger the formation of the gel for the present system of binaphtol derivatives in solution was that of stimulation with ultrasounds. Ultrasounds are frequently used in Applied Chemistry as an effective way to create dispersions of partially immiscible solutions and to disrupt weakly bound ordered structures, such as membranes and

small aggregates. However, as reported by some authors in recent years, sonication can also produce the opposite effect, inducing the aggregation of organic molecules, particularly in the form of gels.

The phenomenon of gelation is often achieved experimentally by simply increasing the temperature of a solution containing gelator molecules above the critical temperature, *Tg*. This is the temperature at which the system spontaneously forms a network of fibres leading to a sudden increase in viscosity, a macroscopic manifestation of the gel formation. Gelation has therefore the characteristic of a thermally activated process.

There are two fundamental factors for the gel formation. On one hand, gelators need to have a distinctive binding configuration allowing for a continuous structure, such as a surface or a fibre, to be formed. Secondly, it is extremely important to take into account the interaction of gelator molecules with the solvent. The solubility of gelators, indeed, is a crucial factor because it influences the properties of the mixing and the interfacial energy. For example, a too low solubility of gelator molecules would indeed favour phase separation in the early stages of the solution, clamping the molecules together and thus providing an obstacle to formation. Moreover, the solubility inherently influences the stability of aggregates by affecting the energy involved in the formation of the sol/gel interface.

Both the aforementioned factors are clearly affected by the temperature. Since energy barriers are more easily crossed at higher temperature, the exploration of the phase space for systems characterised by a collective ordering is greatly enhanced, thus allowing them to reach relative minima that otherwise would be rarely observed. Moreover, under such condition, the temperature inevitably influences the solubility parameters, by either increasing or decreasing the mixing.

Despite a number of examples in which ultrasound-induced gelation was reported in recent years, the understanding of the interactions between LMOG leading to gelation and the role of sonication as viable method to trigger their aggregation have not been yet made clear [64]. Some authors [65, 66] seem to agree on the fact that sonication introduces sufficient energy into the system such that the collective aggregation process can find a minimum that is distinctively different from the minimum achieved by thermal activation. Some experiments were also able to demonstrate that the microstructure of a gel formed by sonication can be different from the gel obtained by a temperature increase above *Tg*. Together with the microstructure, also the properties of the aggregate can vary substantially. For instance, it has been reported [59] that the gel formed by particular cholesterol-based molecules displays remarkably different wettability depending on which method is used.

Other authors [58, 60, 59] have specifically addressed this point by making the hypothesis that gelation can be achieved by sonication if a LMOG molecule can undergo a conformational transition that acts as a seed to the gel formation, similarly to what prion proteins do. The new conformation, indeed, is believed to initiate the aggregation process by promoting the transition of other molecules to the new configuration, thus leading to a network of fibres. Although no further insight into the mechanism has been provided concerning the gelator's transition, the sonication of a solution introduces the conditions for cavitation: sound waves propagating in a fluid induce large fluctuations of pressure which can lead to the formation of micro-bubbles. These bubbles, though, do not last very long and soon collapse catastrophically: the rapid implosion of bubbles can produce a local temperature of above 1000 K and pressure in the range of 10-100 *kbar*. It is well know that these conditions are sufficient to trigger a chemical reaction: cavitation is, indeed, a well known cause of erosion in metals and much attention is devoted to avoiding its effect in engines and propellers. On the other hand, cavitation might be exploited for its potential application in the synthesis of metal-organic compounds and nucleation of particles. In recent years, the use of ultrasounds was demonstrated to be particularly promising in the field of mechano-chemistry [67]. The application of local forces through sonication is able to activate a broad series of reactions involving chemical groups called *mechanophores,* whose reactivity is sensible to the application of stresses. The reaction path may follow different mechanisms with respect to thermally activated process, since the applied forces decrease the energy barriers, thus affecting the reaction rate coefficient according to the following relation posited by transition state theory:

$$k = \kappa \, exp\left[- \left(\Delta E^{\dagger} - F \Delta \xi \right) / k_b T \right] \tag{2.4}$$

where κ is some nontrivial kinetic prefactor, $\Delta E^{\dagger}$is the original activation energy and $F\Delta\xi$ is the work done on the system by a force F along a path ξ. It is then clear that, if sonication and the extreme conditions experienced locally by the system can actively cause chemical reactions to take place by breaking bonds, the same forces can induce conformational transitions for particular molecules. Generally speaking, the energy barriers involved in the transition between different configurations are low enough to allow the whole configurational space to be easily accessible: thermal energy should be low enough to overcome these barriers and connect all the points in Potential Energy Surface (PES). The peculiar characteristic of cavitation, though, is the local character of the stresses introduced while the overall temperature is not substantially affected thus preventing a whole series of reaction from taking place simultaneously. It has indeed been demonstrated that ultrasounds/cavitation have different effect on molecules of different size [68]. This is the case for polymers: particular bonds can be broken by ultrasounds depending on the length of the polymer chains.

Various authors have independently formulated their hypothesis for the gelation mechanism on the basis of the ability of ultrasounds to induce a conformational transition of the LMOG. Naota and Koori [58] first reported about sonication-induced gelation for a particular chiral metal-organic compound. In particular the gel formation was proved to be:

- *stereo-selective*: gelation was achieved only by a heterochiral composition of the solution.
- *reversible*: heating the gel above Tg leads one to recover the solution (this might be a sign of directional interactions like $\pi - \pi$ stackings, which are expected to be more sensitive to temperature changes).
- *mildly sensitive to solvent polarity*: the gelation did not seem to be greatly influenced by the type of solvent used, as far as its polarity was concerned.

Naota *et al.* partially justified the previous finding by identifying one possible conformation of the molecule as responsible for the repetitive aggregation in fibres through $\pi - \pi$ stackings of aromatic rings. Particularly, this explanation seems to address the chiral selectivity of the gelation process that was reported to take place thanks to the interlocking of racemates, which was proposed to be possible for homochiral compounds. Analogously, Yao *et al.* [60] reported about a particular amino-acidic molecule that gels upon sonication and proposed a mechanism relying on the presence of a conformational transition of the molecules promoting the morphological transformation from spherical aggregates to fibres (thus leading to the formation of a network). Another cholesterol-based compound was reported to be gelating upon sonication through conformational change also by Wu *et al.* [59]. Interestingly, slight modifications of the molecular structure were proved to form morphologically different gels (that can be reversible and be formed also upon heating) or even prevent gelation. Remarkable different wettabilities were also measured for gels depending on how they were formed (temperature or sonication).

2.4 Preliminary computational studies

The works cited in the last part of the previous section report some similarities with the experimental findings for binaphtol molecules. In particular, the recurrent hypothesis of gel formation based on the existence of a defined conformational transition induced by sonication recalls the existence of the two conformations in fig. 2.2. In addition, the sensitivity exhibited by the system to the chiral composition of the solute (i.e. R,R vs S,S) allow one to draw a parallel with the findings of Naota. Interestingly, these two factors seem to point towards a unique direction of investigation. The very basis for the gel formation seems to be dependent on a precise

aggregation motif that is strictly related to the existence of a small set of distinct binding configurations, that can be steadily formed only between molecules of the right conformation and with the right chirality. Only through these binding configurations can the gelator molecules clamp in a repetitive fashion, allowing for the formation of supramolecular aggregates. Any other binding configuration, any other stereoisomer or any modification to the native structure is expected to destabilise aggregation.

The micrographies reported in section 2.2.2 allow one to describe with a sufficient level of detail the main features of the different aggregates found in the gel. It was possible indeed to extract some information relative to their dimension and to start drawing a preliminary hypothesis. What is still missing though, even in the images at the highest resolution, is an explanation for the formation of aggregates that is based on the features of the interactions between molecules resolved at the atomistic level. This work aims to fill this gap through the aid of both classical MD and Monte Carlo sampling.

Even though classical potentials are intrinsically much simpler and less precise than most ab initio methods, they are preferred in this case for a number of reasons. Firstly, the size of the molecules combined with the complexity of the configurational space, even just between pairs, put a severe limit to the applicability of highly accurate methods requiring a demanding computational load: faster methods for energy calculations are, therefore, preferred. Secondly, it is well known that DFT, even with a careful choice of the exchange and correlation potential, not always guarantee the expected accuracy in the treatment of vdW interactions correctly: for example, it was demonstrated that $ch - \pi$ interactions, that are likely to be significant for the formation of stable binding configurations, are significantly underestimated for benzenes and naphtalene [69]. Finally, since binaphtol molecules are mainly formed by planar aromatic groups, they can be considered to have a rather rigid structure: the presence of favourable binding configurations very likely arise from a geometric fit that optimises the interactions between nearly rigid objects. For this reason the following study will mainly make use of classical potential, while DFT calculations will be mainly used for comparison.

The aforementioned limitations of DFT, together with a difficult, if not impossible, retrieval of reference experimental values makes the evaluation of the accuracy of a given *ff* in the parametrisation of the molecules extremely complicated. For simplicity, throughout all this study, a unique *ff* will be used. In particular, the GAFF force field [14] was chosen as the reference potential because of its well-documented transferability (see Chapter 1 for a detailed description of the GAFF force field). We decided not to adopt the correction potential for aromatic carbons that will be described in chapter 3, because the binaphtol molecules considered here are entirely formed by aromatic groups while our correction was mainly aimed to fit the

energy profile for the parallel orientation of benzene in a otherwise heterogeneous environment.

2.4.1 The two conformers

As mentioned in section 2.1.1, the binaphtol derivatives can assume two main distinct conformations. Depending on the relative rotation of the naphtyl groups with respect to the benzene rings, a *cis-* and a *trans-* conformers are identified (*folded* and *extended,* respectively). In the *cis-* conformer, two naphtyl groups are oriented towards each other, presumably establishing $\pi - \pi$ stacking interactions between themselves. In the *trans-* configuration the same groups are placed on opposite sides of the molecule, thus $\pi - \pi$ stacking is preventedmore. This observation supported the hypothesis that the folded configuration might be energetically favoured over the extended, unless other factors counterbalance the contribution given by the intramolecular dispersion interaction. In order to check whether this was the case, a minimisation of the potential energy defined by the GAFF *ff* was performed in the gas phase for both structures by means of a conjugate gradient algorithm. Consistently with the observation made ealier, the folded configuration was found to be energetically more stable than the extended one: an energy difference of about 3.76 $kcal/mol$ was reported; in particular this value corresponds roughly to half of the binding energy calculated with the same method for naphatalene dimers. In order to assess this result obtained with a classical potential, a geometry optimisation was also carried out in DFT using the PBE-GGA exchange and correlation potential. The relaxation of the minima found at first presented some differences. First of all, the folded conformer was not found to have a lower energy; on the contrary, the extended conformation was found to be 0.50 $kcal/mol$ more stable than the folded conformer. This result is consistent with the value calculated by Bonifazi et al. (article in preparation, see [70] for reference) that reported a slightly lower energy difference equal to 0.37 $kcal/mol$ (B3 LYP 6-31G** level).

The reverted energy stability of the conformers is reflected in the analysis of relaxed molecular structures. The minima found for the folded and the extended conformers are shown in fig. 2.9 a and b, respectively (the classical minima are rendered in blue, while the DFT ones are displayed in red). The comparison of the structures reveals a significant difference for the folded conformer, for which a RMSD of about 0.81 Å was calculated. This value is explained by a different rotation of two lateral naphtol groups, which is more pronounced for the classical case, possibly due to the attractive aromatic interaction. Such a difference can be expressed by the values of the dihedral angles formed between the atoms connecting each naphtyl group and the benzene rings, in correspondence of the oxygen atoms. As it can be seen in table 2.3, the dihedrals differ by about 10° and 5° for the two dihedrals reported (the other two are equivalent

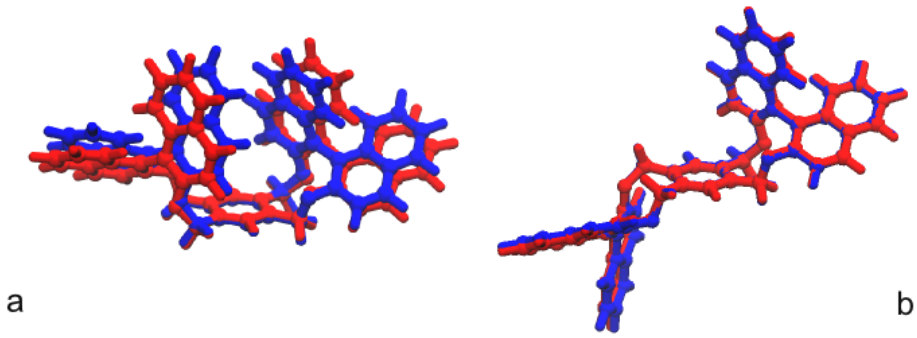


Figure 2.9: Minimised structures for the folded (a) and the extended (b) configuration; the minima corresponding to the classical potential are shown in blue, the DFT minima are shown in red.

for symmetry). As a result, the two facing naphtyl groups are found at a distance, measured between the centre of masses, of about 4.2 Å; in the DFT case, the naphtyl group are, instead, almost parallel, separated by a larger distance, corresponding to 6.2 Å. By contrast, the two methods display a much better agreement for the extended conformers. The matching of the two minima is indeed good and was confirmed by a RMSD value of 0.11 Å; furthermore, the dihedrals obtained correspond much better than the previous case, with a difference of only about 1.5° and 2°.

		Dihedral 1	Dihedral 2
Folded	DFT	-61.7	-144.2
	Classical	-51.3	-139.1
Extended	DFT	-67.9	-148.6
	Classical	-66.4	-146.5

Table 2.3: Comparison of the dihedral angles for the minima found for the two configurations for the classical potential and DFT.

2.4.2 The transition energy barrier

Despite the contradictory results concerning the minima found for the classical potential and for DFT, it is important to understand whether the transition between the two conformers is possible. This is particularly significant for the consideration reported in section 2.3 about the role of a conformational switch as a possible mechanism triggering gelation. In order to establish this, it is necessary to find the reaction path that connects the two minima in the PES. A preliminary observation of the structures leads one to assume that the process responsible

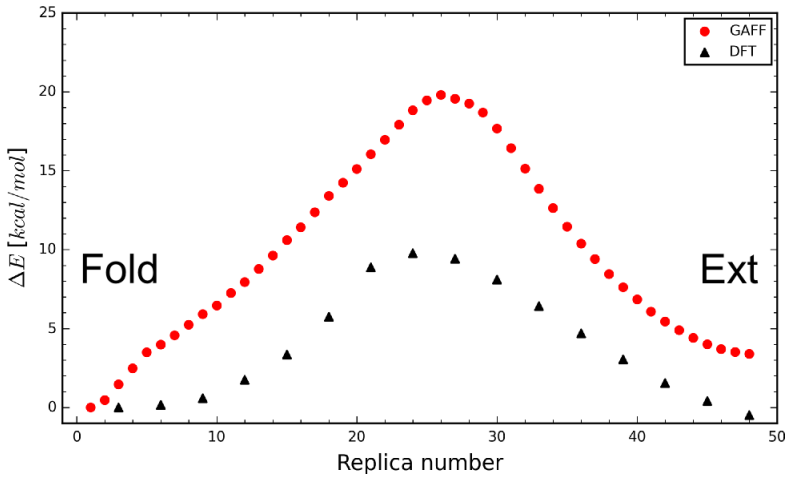


Figure 2.10: The potential energy is plotted as a function of the replica number for the NEB calculation performed via classical potential (red circles) and DFT (black triangles). A higher classical activation energy was calculated for the extended-to-folded transition, closely matching the value reported for the experimental results.

for the transition from one structure to the other is the rotation of one of the two binaphtol groups around the central benzene ring. A series of NEB calculations were then performed both with the classical potential and with DFT to estimate the energy barrier involved in this rotation. For the first case, a series of 48 replicas were simulated with an implicit solvation model (pairwise generalized Born model [71]) till equilibration, and then subsequently annealed to 0 K; for the DFT case, 16 replicas were used. An intermediate configuration (derived by the classical simulations) was set as climbing image in order to help convergence.

The results reported in fig. 2.10 show the energy profiles for the two methods as a function of the replica number. As expected, the activation energies are noticeably different: for the classical NEB, a value of 16.41 *kcal/mol* was found, while a lower barrier of about 9.78 *kcal/mol* was calculated with DFT. The analysis of the trajectories for the two paths revealed that the energy barrier in correspondence of the saddle point is determined by a wide variation of the aforementioned dihedral angles, combined with the rotation of both the hydrogens of the two sp3 carbons. The experimental estimation of the energy barrier was also carried out by comparing the intensity of the NMR spectra of a solution of binaphtols containing both conformers at different temperatures yielding a value of 15.8 *kcal/mol*. This activation energy lies in between the two values found with DFT calculations and the classical potential. In particular, it is significantly closer to the classical activation energy, which is just 0.6 *kcal/mol* higher. On the contrary, the energy barrier calculated with DFT is about 40% lower than the experimental one.

The classical activation energy was calculated with respect to the higher energy minima i.e. the extended configuration - that was demonstrated to closely match the relaxed DFT structure.

However, since the folded conformer did not display an analogous agreement, the comparison with the experimental results has to be considered carefully: the reported differences for the dihedral angles suggest, indeed, that angle-dependent contributions in the potential energy, while not describing accurately the folded structure, might not lead to a precise estimation of the energy differences along the reaction path. However, two series of NEB calculations set an energy interval consistent with the experimental findings, within which the real value would be likely found. The importance of this calculation, indeed, does not rely in the identification of a precise and defined activation energy, but is intended to verify whether the transition between the two isomers was not prevented by steric hindrance. Importantly, the activation energy is sufficiently low to consider the conformational switch a process that can be activated by a energy in the range of kT, thus not requiring additional external contribution such as sonication.

2.5 Solvation energies for different solvents

Table 2.1 lists the full set of investigated solvents, together with their dipole moments and the resulting behaviour upon sonication. Previous studies on different aromatic systems [58] highlighted how the aggregation process was not found to depend much on the nature of the solvent. Here, the experimental findings for binaphtol molecules show how, among the whole series of tested solvent, only for methanol and cyclohexane the process was observed, while, in all the other cases, the solutes either remained in solution or precipitated, for analogous quantities of gelators. These findings are consistent with the fact that gelation arises from an interplay between the solvation energy and the cohesive energy of the aggregates. Unfavourable solvent interactions would likely increase the tendency of gelators to segregate, while favourable ones are expected to increase the mixing of molecules, thus preventing the formation of fibres. These two scenarios are consistent with part of the findings reported in table 2.1. The binaphtol derivatives are indeed mildly polar molecules characterised by a dipole moment that is null for the extended configuration, while a value of 3.44 D was calculated for the folded one. Highly polar solvents, such as nitromethane and acetonitrile, would result in a scarce solubility of the solutes, which in fact were found to precipitate. On the other hand, the higher the affinity of the solutes with the solvent, the more stable is the solution, which is not, in fact, altered by sonication as revealed by many entries in table 2.1.

Despite leading to similar results, methanol and cyclohexane are considerably different molecules. The first is the shorter alkyl alcohol with a dipole moment equal to 1.70 D, while the second is a non-polar cyclic alkane often used as an organic solvent. To study whether their polarity is a relevant parameter for the system, the calculation of the solvation energy for both the conform-

ers was carried out for the two solvents. Additionally, a third solvent molecule was considered for comparison: tetrahydrofuran was chosen since it has a dipole moment similar to that of methanol, while not showing comparable gelation capabilities.

The calculation of solvation energies presented some challenges. A commonly used approached relies on thermodynamic integration: soft-core potentials are used to integrate the potential energy variation from the initial to the final state. This method allows one to assess the solvation energy differences for amino acids and small substituents [72, 73, 74], while it cannot certainly be used for bigger molecules. On the other hand, the implicit models such as those described in chapter 1, were considered too coarse for this case. Another strategy was therefore devised. Firstly, the energy of each conformer alone was sampled in vacuum by means of Langevin dynamics (collision frequency 1 ps^{-1}) at 300 K; then the same molecules was equilibrated in solvent for a total of 10 *ns*. Trajectories of the solute molecule were then extracted for the solvated systems and the energy was calculated again in gas phase. The comparison for the average energies led us to evaluate the energy difference for the solute molecule induced by solvation. The corresponding values are reported in table 2.4.

As expected, the calculated solvation energy for the folded conformer in methanol reported the lowest value, about -5.4 *kcal/mol*. The solvation energy for the same conformer in a non-polar medium was instead -3.1 *kcal/mol*, thus showing a significant reduction of about 2.7 *kcal/mol* with respect to the polar medium. Differently, the low polarity of the extended conformer yielded similar solvation energies in both solvents, equal to -2.7 *kcal/mol* and -2.6 *kcal/mol* in cyclohexane and methanol, respectively. The same calculations were then performed for tetrahydrofuran, yielding analogous results to those found for methanol. Also in this case, the higher polarity of this solvent favoured the interaction with folded conformer, while, for the extended one a higher energy was calculated, equal to about -2.7 *kcal/mol*. Evidently, what emerges from this comparison is that the polarity of the solvent molecules and the solvation energies alone are not sufficient to justify the occurrence of gelation and, conversely, the missing gelation in tetrahydrofuran. This allows to hypothesise that the contribution of solvation is expected to be relevant in specific binding configurations and in the creation of the interface with larger aggregates.

	ΔE_{sol} cyc $[kcal/mol]$	ΔE_{sol} moh $[kcal/mol]$	ΔE_{sol} thf $[kcal/mol]$
R,R extended	-2.7	-2.6	-2.7
R,R folded	-3.1	-5.4	-5.3

Table 2.4: Solvation energies for the two R,R conformers in cyclohexane (2nd column), in methanol (3rd column) and tetrahydrofuran (4th column); the standard error of the estimated means is equal to 0.2 *kcal/mol*, confirming the statistical significance of these averages.

2.6 Binding of binaphtol molecules

	ΔE $[kcal/mol]$
R,R-ext. R,R-ext.	-14.3±1.7
R,R-folded R,R-ext.	-10.5±1.2
R,R-folded R,R-folded	-8.7±1.4
S,S-ext. R,R-ext.	-9.5±1.3
S,S-ext. R,R-folded	-10.0±0.9
S,S-fold R,R-folded	-7.0±1.1

Table 2.5: Average binding energies for homochiral and heterochiral pairs of conformers in solutions; all possible pairs are considered, excluding the redundat cases which are redundant for symmetry (e.g S,S pairs); error bars referes to the standard errors.

In the previous section, the solvation energies of the gelators were presented and discussed for three different solvents. In order to elucidate the rationale behind the occurrence of gelation and particularly its chiral dependence, a detailed description of the interactions at a molecular level is required. Before even attempting to study the aggregation of multiple molecules in a solvated system, it is convenient to focus the attention on the interaction of pairs of molecules. This allows one to restrict the problem to finding a set of binding configurations that may support the existence of a supramolecular order. Since the gelation capability of the molecules was observed to be an increasing function of the enantiomeric excess, three pairs of homochiral binaphtols derivatives (R,R-R,R) were studied first, in an attempt to account for the possible combinations arising from the conformers reported in fig. 2.9. This required us to consider the following pairs: extended - extended, folded - extended and folded - folded. For comparison, the interaction between heterochiral molecules was considered by studying three additional systems consisting of the same conformers as before but with opposite chirality (R,R-S,S) .

The six pairs of molecules where then solvated in a box of methanol and 80 *ns* of dynamics

were carried out in the NVT ensemble at 300 K with a Langevin thermostat. During this simulation time, the radial distribution functions were monitored and showed to converge nicely (the logarithm of the probability distribution was used for comparison with the data in table 2.5). From the equilibrium trajectories 250 configurations were selected in order to calculate the average binding energies. Instead of randomly picking the frames, the distance between the centres of mass of the molecules was used to determine whether the configuration pertained to a bound state: only the frames for which the aforementioned distance was lower than 20 Å (approximately equal to the size of a binaphtol) were considered. The calculated energy averages are reported in table 2.5.

The data in table 2.5 indicate that the R,R extended - extended pair has by far the most favourable interaction energy, for which a value of -14.3 *kcal/mol* was determined; on the contrary, the folded conformers bind less strongly, having an average binding energy of -8.7 *kcal/mol*; for the folded and an extended pair an intermediate value was found, equal to -10.5 *kcal/mol*. Similar results for the heterochiral pairs were found in the folded - folded and folded - extended systems, while a significant difference was calculated for the extended - extended heterochiral case: the calculated average binding energy is equal to -9.5 *kcal/mol*, which is roughly 5 *kcal/mol* less than the one calculated for the homochiral dimer. This finding is particularly important because it provides the first quantitative evidence that shows how chirality plays a important role in establishing the interaction strength of dimers.

The data presented in table 2.5 tells us little about the binding geometry of binaphtol dimers. To attempt the resolution of the binding geometry, it is necessary to analyse the configurations to spot possible structural features that can support the energy differences found and particularly the effect of chirality. To do so, the configurations extracted from the trajectories were further minimised with the same potential by using an implicit solvation model (see section 1.3.2 for details); the final structures are reported in fig. 2.12.

The binding configurations obtained from the series of minimisation show in nearly all cases

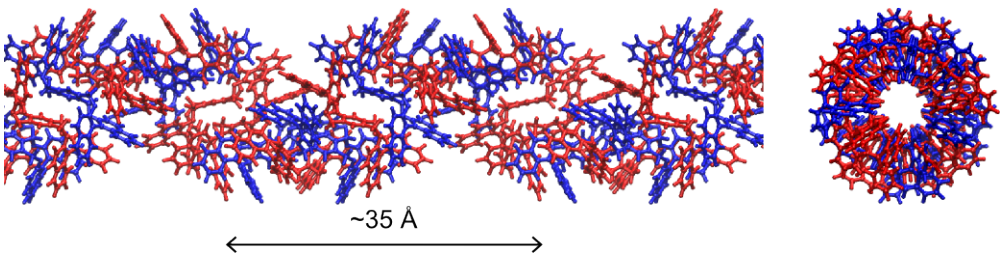


Figure 2.11: One of the dimer configurations of the R,R extended isomer (fig. 2.12 a) can be arranged into a helix.

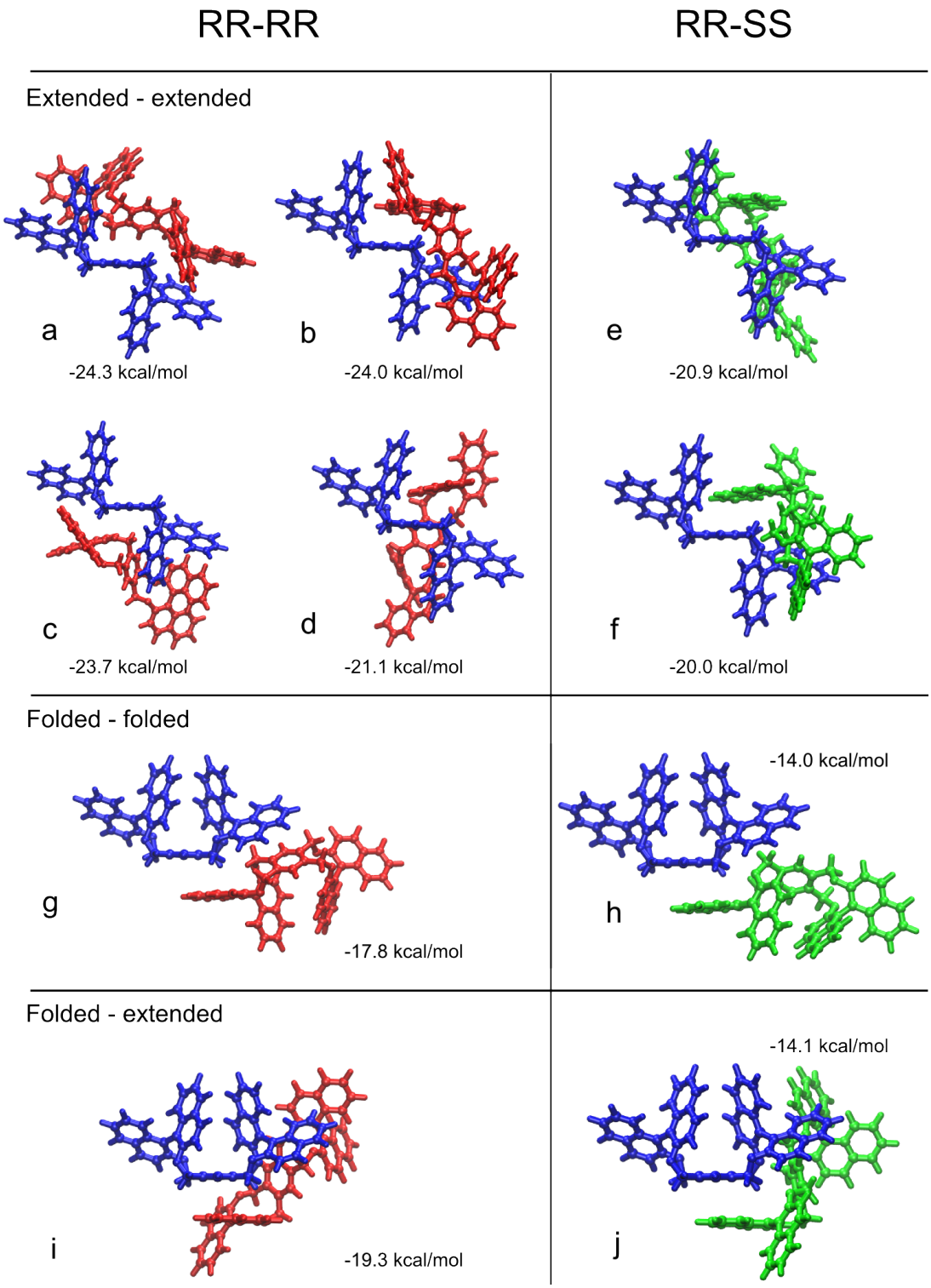


Figure 2.12: Most relevant minima for binaphtol pairs: the extended-extended, folded-extended and folded-folded structures are reported for the R,R atropoisomer pairs (left) and the heterochiral pair R,R-S,S (right).

how the molecules interact mainly through the formation of $\pi - \pi$ stacking between one naphtyl groups and the benzene ring; this was observed for all pairs regardless of their chirality. Among the whole series of configurations, one is of particular interest. In the minimum represented in fig. 2.12 a, two extended molecules are paired along their length with a slight twist. The calculated binding energy is -24.3 *kcal/mol,* which is the lowest found between all pairs. A careful examination of this geometry revealed a particular feature: the consecutive stacking of more extended conformers leads to the formation of a helix. This finding was believed to be important since it has been shown how the self-assembling of small chiral peptides into helical tapes, leads to the formation of larger fibres and gels [75]. A helix formed by 10 monomers was created and annealed in the gas phase using periodic boundary conditions: a stable structure was obtained with a pitch of about $\sim$35 Å (see fig. 2.11). However, when tested in methanol at 300 K, the helix soon lost its original shape, thus indicating that this structure is not expected to be found in real gels.

2.7 The dimers

In the previous section, the interaction arising between binaphtol pairs generated evidence for how the extended conformers interact more strongly when considering homochiral pairing. In fact, the binding energy of heterochiral extended pairs was found to be reduced by $\sim$20% with respect to the homochiral one. To further the study, a model that allows us to monitor the formation of aggregates was necessary to observe whether the binding configurations detected lead to a directional/repetitive arrangement of molecules. Two bulk models were then built, consisting of 20 single molecules solvated in methanol. Given the stronger binding energies reported between pairs of extended conformers, a first system was composed entirely by the extended conformer while in the second one also the folded conformer was included (10 extended and 10 folded). Each of the individual run was carried out for a total of 200 *ns* at 300 K with a Langevin thermostat (preceded by density equilibration). The analysis of the trajectories did not reveal a clear formation of aggregates: molecules did clump together but no stable and extended structure was observed. Many of the binding configurations of fig. 2.12 were detected but it was not possible to distinguish relevant patterns suggesting ordered aggregation features. Starting the simulations from pre-assembled clusters was also conceived to be a good strategy to overcome unnecessary long nucleation times, but this did not showed to lead to better results, since selected starting aggregates did not evolve to supramolecular structures but , on the contraty, showed to loose soon the original cohesion. The same simulations were conducted also at different temperatures to overcome the diffusional barrier and thus increasing the sampling of

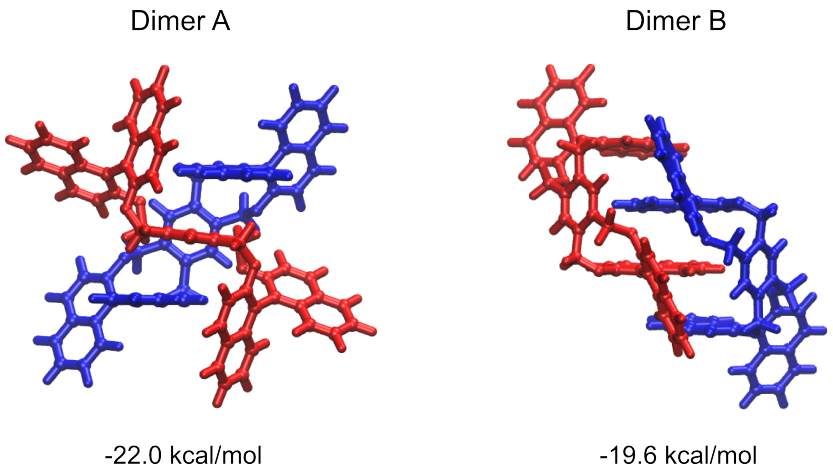


Figure 2.13: Two dimers discovered from the analysis of bulk model trajectories: an extended conformer inserts between the two facing naphtyl group of the folded isomer (left); two folded conformers bind through the formation of interlaced $\pi - \pi$ stackings (right).

the phase space, without yielding significant results. Rather, the collective characteristic of the process indicated that the system might need to be studied through the application of a biasing potential to explore more accurately the phase space (metadynamics or REM). Despite being an ideal method, it was not considered viable due to the difficulty of determining representative collective variables.

The apparent lack of any significant results obtained from these bulk models, nonetheless allowed for the identification of two interesting binding structures that were not observed before. In one case, an extended conformer was found trapped between the facing naphtyl group belonging to a folded conformer by forming a stable dimer (*dimer A*), that remained bound over the course of 15 *ns* of the dynamics. This stability was confirmed by the calculation of the binding energy in the gas phase, corresponding to a value of -22.0 *kcal/mol*. Interestingly, the stability of the dimer is further supported by the formation of a clear interaction pattern: firstly, the distance between the plane containing aromatic groups of 3.5 Å confirmed the formation of $\pi-\pi$ stackings [69]; moreover a series of $ch-\pi$ interactions was detected between the two benzenes (resembling the notorious T-shape minima of benzene [76, 77]) and between the naphtyl groups. A second dimer (*dimer B*) was formed instead by two folded conformers interlocked by $\pi - \pi$ stackings between the facing naphtyl groups. The resulting energy of this dimer was slightly lower than the previous case by roughly 2.5 *kcal/mol*, possibly due to a lower number of $ch-\pi$ interactions formed.

The binding energy of *dimerA* alone is sufficient to foresee its importance in the aggregation of the present molecules. Undoubtedly, the energies reported in fig. 2.12 for the extended/extended

pairs are comparable and in some cases even higher, but this dimer presents other features that are unique. First of all, the alignment of aromatic groups and formation of $ch-\pi$ bonds provide a strong argument for the stability of this configuration. Their energy contribution is not accurately reproduced at the DFT level, thus the relaxation of these structures was not considered of much help; on the other hand, the limited directional characteristic of the classical potential used allows one to hypothesise an even higher stabilization effect when compared to the other configurations reported in fig. 2.12. More importantly, the binding energy for the same dimer structure formed by two molecules of opposite chirality was found to be far less stable. The very same dimer formed by S,S and R,R atropoisomers was indeed repeatedly minimised analogously to what was done for the homochiral dimer, yielding a minimum binding energy of -11.0 *kcal/mol*, 50% less favourable than the homochiral one. The reason for this difference is likely found in the less convenient steric arrangement induced by the different chirality. It is clear, indeed from the structure of the RR-SS dimer reported in fig. 2.14 that the parallel alignment of the aromatic rings is lost; as a consequence, the formation of $ch-\pi$ bonds between the naphtol moieties is significantly hindered. A similar observation can be made for *dimer B,* for which the mirroring atropoisomer could not even yield the same neat stacking of naphtalene groups for a comparable interlocking was not achievable (no calculations were conducted in this case, since no starting guess was possible).

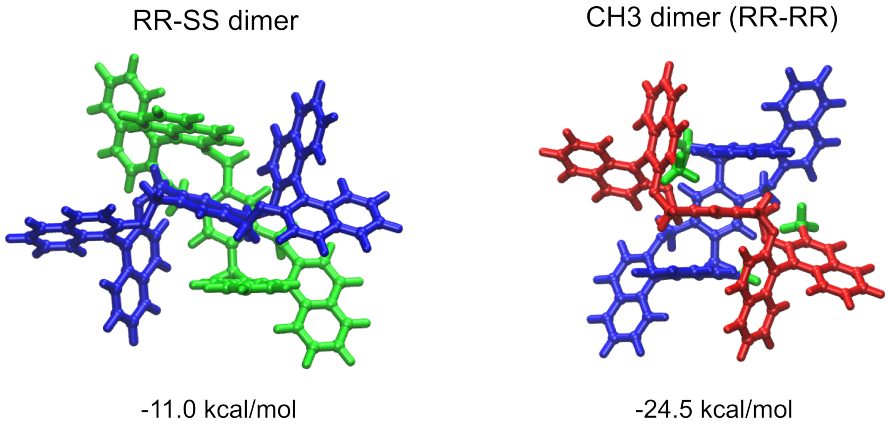


Figure 2.14: The binding energy of dimer A is greatly reduced for heterochiral pairs because of the increased steric hindrance of the binaphtyl groups of opposite chirality (left); a more favourable interaction energy was instead calculated for methylated binaphtols (right); the methyl groups are rendered in green for clarity.

The individuation of these dimers is important also for another reason. One of the indication provided by the experiments concerned the determination of the hydrodynamic radius of the smaller particle extracted from the gel's fibres, which is not compatible with the size of the binaphtol molecules alone; on the contrary it is coincident with the size of the dimer. Considering

the agreement with the measurements, the homochirality intrinsic characteristic of the dimers and its lower energy with respect to most other conformations, the explanation for the chiral-sensitive gel formation might lie in the interactions of these dimers only.

To gather further information that could improve our understanding of the gelation process, a particular modification of the original naphtol moiety was introduced to check whether the replacement of an hydrogen with a bigger group i.e. a methyl - could have any effect on the gel formation. Importantly, this was exactly the case: the addition of a methyl substituent to the the naphtyl groups (rendered in green for clarity) in the position shown in fig. 2.14 was observed in experiments to inhibit completely the aggregation, even for an enantiomeric pure composition. Initially the methyl group was added with the idea that it could have prevented the formation of the dimer A. The calculation revealed that this was not the case, but that, on the contrary, the resulting binding energy of the dimer was lowered to a value of -24.5 *kcal/mol* (calculated in implicit solvation as before).

In spite of the fact that the dimer formation was observed to proceed through lateral interlocking of molecules, the analysis of configuration d (cf. fig. 2.12) suggested that dimer A (fig. 2.13) could be potentially formed upon the extended-to-folded transition of one naphtol belonging to the former extended-extended *d* pair type. This hypothesis was investigated to verify whether the energy barrier calculated in section 2.4.2 could somehow be lowered by the presence of a neighbouring binaphtol. Classical NEB calculations analogous to those presented in section 2.4.2 for the extended-to-folded transition were conducted without producing supporting evidences: the energy barrier did not, indeed, decrease. The same calculation was carried on also for the extended-to-folded transition of methylated binaphtols, to check whether the methyl groups could cause a significant steric impediment to the rotation of the lateral binaphtyl groups, thus increasing the energy barrier for the formation of this dimer. Also in this case the energy was found to be about 18 *kcal/mol*, thus the same order of magnitude as for the non-methylated molecules.

In conclusion, the dimer A was not shown to form with the same effectiveness for a heterochiral pair, for which a binding energy 50% lower was calculated with respect to the homochiral one. This is an important piece of evidence that clearly indicates how chirality affects the energetics of the dimer A. This finding, together with the consideration reported earlier on the compatibility of the hydrodynamic radius of particles evidenced experimentally with the dimer, provides strong reason to further the study in the direction of understanding whether gel formation might be rationalised by the interlocking of dimers.

2.8 10-dimers clusters

The micrographies in fig. 2.6 show spherical aggregates and tubular fibres characterised by a porosity almost certainly caused by solvent evaporation. These cavities allow resolution of the fine structure of the gel, which appears to be composed by quasi-spherical molecular sub-clusters. The size of these particles is rather homogeneous (see left of fig. 2.6) and corresponds to about 5-10 nm in diameter. In order to check whether the gelation process could be rationalised in terms of the aggregation of spherical small clusters, we decided to investigate this possibility by setting up a series of molecular simulations with the aim of establishing the existence of a clear energy dependence based on the clusters' chiral composition. These were also targeted to identifying possible aggregation patterns that could rationalise the formation of larger aggregates and, in addition, how these would be modified by the presence of different atropoisomers.

The first thing to consider was the determination of a reference size for the quasi-spherical sub-cluster. The observed experimental cluster size of about 10 nm is, in fact, too large for addressing clusters' coalescence via full atomistic molecular models, unless some coarse grained model was available. Since this last possibility is excluded by the fact that no such models were made available in the literature (apart from some similar strategies which consider molecules as rigid objects [78]), it was necessary to consider the problem from another perspective.

To tackle this, a viable strategy is to consider a combinatorial approach analogous to the Flory–Huggins theory of solutions. Since no evidence of any molecular sorting effect was reported in experiments, we considered a cluster of N dimers composed by n dimers of one chirality and $(N - n)$ of the opposite. The total number of different clusters that can be generated is found simply by considering the possible combinations of N particles in n sites (the symmetry of the system is not considered in this case, for simplicity). Hence, the cluster count is given by the relation:

$$C_{N,n} = \frac{n!}{N!(N-n)!}. \tag{2.5}$$

The probability for the dimer of one type to be in a cluster is a function of their relative concentrations in solution. This probability can be thus expressed in terms of the fraction p of molecules of one chirality (which is a constant factor). By using a binomial distribution, we can obtain a measure of the probability distribution for a cluster probability distribution as a function of p:

$$P(N,n) = p^n(1-p)^{N-n}C_{N,n}. \tag{2.6}$$

Since gelation was shown to be a function of the solute enantiopurity (varying from perfect gelation for homochiral solutes, to complete solvation for heterochiral ones), an asymmetric

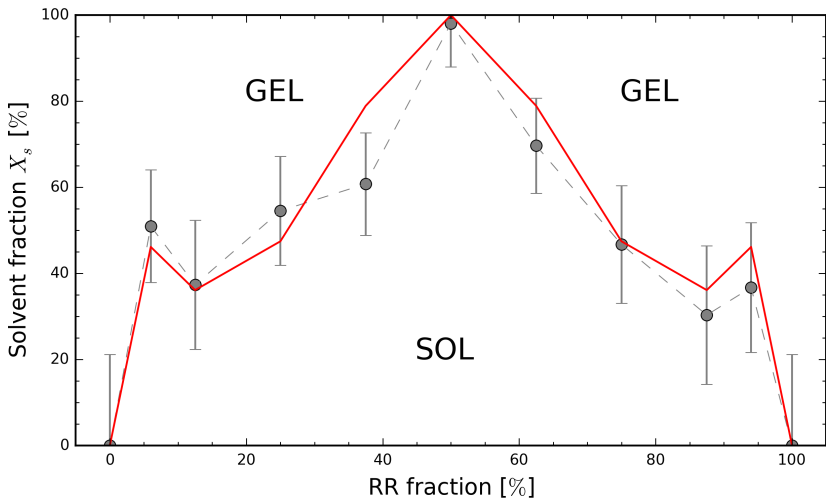


Figure 2.15: The solvent fraction calculated from the relaxation times as a function of the fraction of RR enantiomers (dotted line) is nicely recovered by the combinatorial model.

solution can still form a gel until the relative concentration of the mirror isomer becomes too big, stabilising the solution against gel formation. In the hypothesis that fibres are formed by the agglomeration of small clusters, only those clusters having at limited fraction of binaphtols of opposite chirality will contribute to the formation of the gel, whether or not they are formed. This can be translated into the following expression for the fraction of solvent/gel fraction X_s, as a function of the solute composition

$$X_s(x) = \sum_{n=0}^{N} \sum_{p=0}^{1} P(N,n) g(p). \tag{2.7}$$

$g(p)$ represents a step function which takes into account only clusters within a range of a given composition, it is therefore either equal to 1 or 0. By taking $g(p) = 1$ for $p = min(\frac{n}{N}, 0.25)$ we evidenced a remarkably good agreement with the experimental findings for a cluster of $N = 10$ dimers. The plot in fig. 2.15 shows both the curves derived from the NMR relaxation measurements and the one obtained from our combinatorial model. Interestingly, this inferential approach nicely fits the experimental solvent fraction with remarkable resolution, which captures also the symmetric peaks of X_s measured at R,R fraction equal to 6% and 94%.

A cluster of 10 dimers appears to be sufficiently big to reveal any possible pattern arising from the interactions of the particle contained in it. In particular, the problem of sampling the orientation of molecules is significantly reduced with respect to a larger system. A series of 6 systems composed of R,R binaphtol dimers composed by an increasing number of S,S atropoisomers was studied via MD. All simulations were carried out without the use of periodic

boundary condition, since this analysis is ultimately devoted to the study of isolated aggregates. However, to confine the cluster within a limited region of space and avoid undesired diffusion of molecules far apart, a weak, repulsive spherical potential was applied. This spatial confinement was implemented using the following 9-3 LJ potential, which was applied to particles within a sphere of radius r_c:

$$V_{sph}(r) = \varepsilon \left[\frac{2}{15} \left(\frac{\sigma}{r} \right)^9 + \left(\frac{\sigma}{r} \right)^3 \right] \quad r < r_c. \tag{2.8}$$

The parameters contained in eq. 2.8 were chosen to recover a pure repulsive potential while limiting as much as possible spurious contributions: in this sense, ϵ was set to 0.01 *kcal/mol* in order to reduce influence of the weak attractive term proportional to r^{-3}; a value of 2 Å was chosen for σ, while the cut-off of 2.5 Å was applied. The dynamics of each system was integrated over 170 *ns*, during the first 20 *ns* of which r_c was decreased from an initial value of 100 Å to the final value of 30 Å. From the equilibrated trajectories, 250 cluster configurations were then randomly selected for each system and annealed to 0 K: the energy of the resulting minima was subtracted by the energy of a single dimer in the gas phase and divided by the number of dimers (N); the average energies are reported in fig. 2.16 as a function of the compositions.

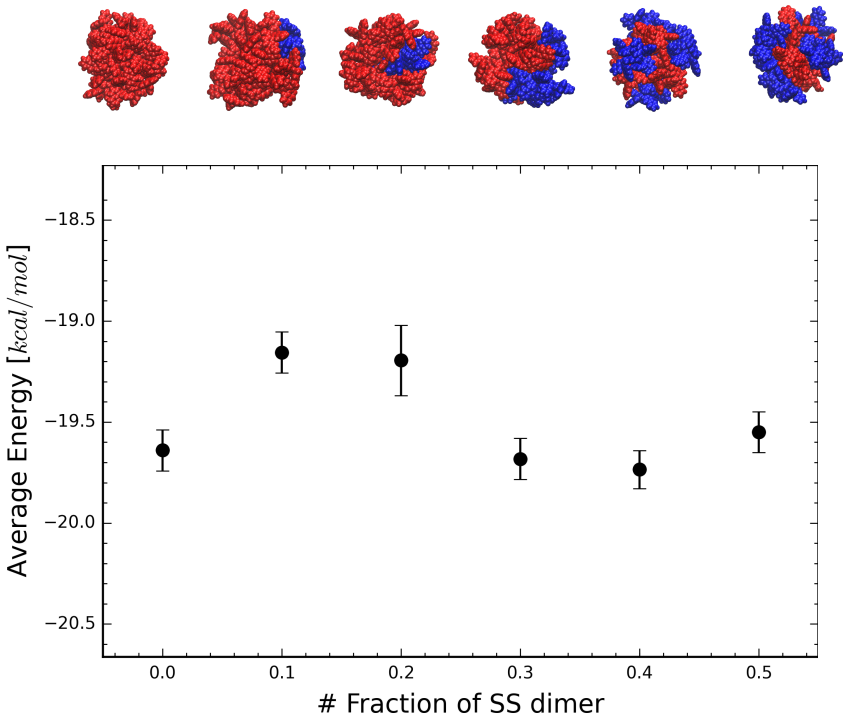


Figure 2.16: Average energies for the 10 dimer clusters of different compositions; the respective clusters are rendered in the figures above the graph: R,R dimers are displayed in red, S,S ones in blue.

Examination of the plot reveals that a large energy dependence is not found for the mean

cluster energies. For the homochiral system, an average interaction energy of -19.7 *kcal/mol* per dimer was found. An increase of about 0.6 *kcal/mol* was observed for the fraction 0.1 and 0.2 corresponding to the presence of 1 and 2 dimers, respectively. This seems to provide an indication that a small fraction of heterochiral components do affect the interaction energy. However, for higher fractions, the average interaction energies is within the same range of the homochiral system. The error bars displayed in the graph are relative to the standard errors, which were calculated to be equal to about 0.2 *kcal/mol*. The distribution of the energies around the mean values, on the other hand, are much wider, about 3 *kcal/mol,* and are not reported since they do not correspond to a physically meaningful statistical distribution; rather, they suggest that the interaction energy of a cluster greatly vary depending on the configuration of the dimers, while no clear minima was found. As a result, these averages do not provide a sufficiently strong evidence to support the effect of chirality on clusters' formation. In particular, the energies reported in fig. 2.16 show that the binding energy between mirroring dimers R,R-S,S might not differ substantially from the energy between dimers of the same chirality. This can be regarded as an argument consistent with missing gel sorting behaviour mentioned earlier: racemates do not segregate and the formation of a gel R and a gel S was not observed. Moreover, the lack of any significant energy difference between different compositions agrees well with the model of eq. 2.7, in which no energy term was accounted for. This is consistent with our hypothesis for which gelation might, in fact, occur as a result of the aggregation of the 10 dimer clusters; we speculate, in fact, that aggregation might be explained by the existence of interlocking configurations that are hindered for heterochiral pairs for steric reasons and can hardly be identified on average.

2.9 Sampling the space of rotations

A thorough characterisation of the pair interaction necessitates scrupulous sampling of the possible orientation with which two molecules come into contact. This procedure is relatively straightforward when dealing with small sized molecules. In this case, both molecular dynamics and Monte Carlo sampling yield satisfactory results. Conversely, when analysing larger, asymmetric molecules, it is much more difficult to sample the whole configurational space. The rotational relaxation of molecules of large size is a much slower process and it is expected to depend on the visco-elastic properties of the fluid where they are immersed in. In the context of gel formation, sonication is expected to play a central role, since it induces an oscillating stress field that was shown to affect rheological properties [79]. However, this problem, essentially a fluid dynamics one, is beyond the scope of this work.

To circumvent this issue, one might consider a system containing more than just two molecules

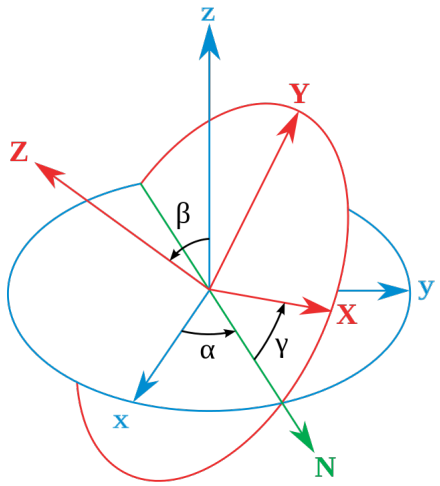


Figure 2.17: Representation of Euler angles describing the rotations between two reference frames (source: http://en.wikipedia.org/wiki/Euler_angles).

to enhance the sampling by simply adding degrees of freedom, as tentatively done for the aforementioned bulk models. In this way, it should be possible to observe the aggregation of molecules but, at the same time, it would be rather complex to discern between energy contributions arising from nearest-neighbour interactions and those related to the supramolecular aggregation. In other words, it would be increasingly difficult to detect any relevant structural information on the binding mechanism other than the trivial ones (if any). This is particularly true when molecules are inherently asymmetric. On the other hand, by limiting the system size to two molecules only, we might easily get trapped in a defined region of the potential energy surface. Alternatively, if the volume of the system is too big, we might end up exploring a small subset of the PES, perhaps far away from the true minima.

The above considerations suggest that an effective strategy to study the interaction of large, asymmetric and relatively rigid molecules is to sample the phase space with an intrinsically statistical approach, instead of collecting statistics from a deterministic sampling. For this reason, we preferred to implement an ad-hoc Monte Carlo sampling algorithm rather than further exploiting MD. In order to do so, the first problem to tackle is how to generate a rotation from a uniform random distribution.

2.9.1 Uniform sampling

The description of the relative orientation of two rigid bodies is equivalent to identifying the rotation matrix that aligns the reference frames of each rigid body. The rotation space in the Euclidean coordinate system, also known as the *SO3* group, can be represented by three numbers

correspondent to the rotation around three axes. A proper choice of the axes, though, is not unique. Euler angles are a very common choice because they allow for an intuitive representation. The scheme reported in fig. 2.18 shows how the initial frame xyz can be aligned to a generic XYZ frame through three consecutive rotations of Euler angles α, β and γ. The corresponding axis of rotation is the z axes of the initial frame, the N axis, which is the intersection of the planes xy and XY (called nodal line), and the Z one.

Euler angles, though, are not always considered the most convenient way of representing rotations. The reason for this is that despite giving a numerical description of any rotation in the three dimensional space, Euler angles are not a covering map and present a topological constraint, also known as the gimbal lock. This problem arises from the fact that there are some points in the space where an infinitesimal rotation cannot be realised by a corresponding variation of the angles. Another common way of representing rotations is based on Euler's rotation theorem which states that any rotation can be reduced to a single rotation around some axis. According to the so called axis-angle representation, the relative orientation of two frames is then reduced to the determination of a normalised vector V in R^3 and the corresponding angle χ of rotation. This vector can be can be written using its projections on the original reference coordinates through the angles θ and ϑ to yield the following:

$$\mathbf{V} = (sin\vartheta cos\theta, sin\vartheta sin\theta, cos\vartheta). \tag{2.9}$$

The rotation matrix can therefore be written by simply deriving the expression of a rotation around V of angle χ, expressed in a compact way as a function of the cross product $V \otimes V$

$$\mathbf{R} = cos\chi\mathbf{I} + sin\chi\mathbf{V} + (1 - cos\chi)\mathbf{V} \otimes \mathbf{V}. \tag{2.10}$$

To correctly sample the rotation space it is not sufficient to sample uniformly the angles, as one might easily realise; it is necessary, in fact, to sample the correct distribution, resulting from the chosen representation. We report here the probability density for both the Euler and the axis-angle representations as in [80]:

$$g(\alpha, \beta, \gamma) = \frac{1}{8\pi^2} sin\gamma, \tag{2.11}$$

$$g(\theta, \vartheta, \chi) = \frac{1}{2\pi^2} sin^2\frac{\chi}{2} sin\theta. \tag{2.12}$$

To generate a random orientation, we need to sample the distribution g. This can be achieved by using random numbers. Given three (or four) random numbers w_i in the interval $[0, 1]$, we

can equate the following integrals,

$$\iiint g(\alpha,\beta,\gamma)d\alpha d\beta d\gamma = \int g(\theta,\vartheta,\chi)d\theta d\vartheta d\gamma = \int dw_1 \int dw_2 \int dw_3. \tag{2.13}$$

This condition corresponds to equating the distribution arising from the probability density g to the product of three independent uniform distributions. By inserting 2.11 and 2.12 in 2.13 and applying the correct integration limits, we get that, for the Eulerian representation, α, β and γ are generated within the correct distribution if the angles w_1 and w_2 are uniformly distributed in the range $[0,\pi]$ and $[0,2\pi]$, while the third angle is obtained as reported in eq. 2.16,

$$\alpha = \pi w_1, \tag{2.14}$$

$$\beta = 2\pi w_2, \tag{2.15}$$

$$\gamma = cos^{-1}(1-2w_3), \tag{2.16}$$

where w_i are uniformly distributed random numbers in the interval $[0,1]$. For the axis-angle representation, the angles ϑ and θ can be calculated as done for the Eulerian β and γ (eq. 2.14 and 2.15, respectively). Since one of the integral in eq. 2.13 is not reversible, the sampling of the remaining angle χ requires an additional random number. Hence, the full set of sampling equations for the axis-angle representation is

$$\theta = 2\pi w_1, \tag{2.17}$$

$$\vartheta = cos^{-1}(1-2w_2), \tag{2.18}$$

$$\chi = \frac{1}{2}(\pi w_3 + cos^{-1}w_4). \tag{2.19}$$

Using one of these two representations is then possible to generate a rotation matrix which can be used to orient a molecule in a random way. In this work, the second method was preferred since it allowed a more intuitive representation that greatly helped debugging. In fig. 2.18 a comparison between a non-uniform (top panels) and a uniform (bottom panels) distribution of the vector $\mathbf{V}$ is displayed: the non-uniform sampling shows a higher density around the poles, while the algorithm implemented - obtained using eq. 2.17 - resulted in a homogeneous scattering of orientations.

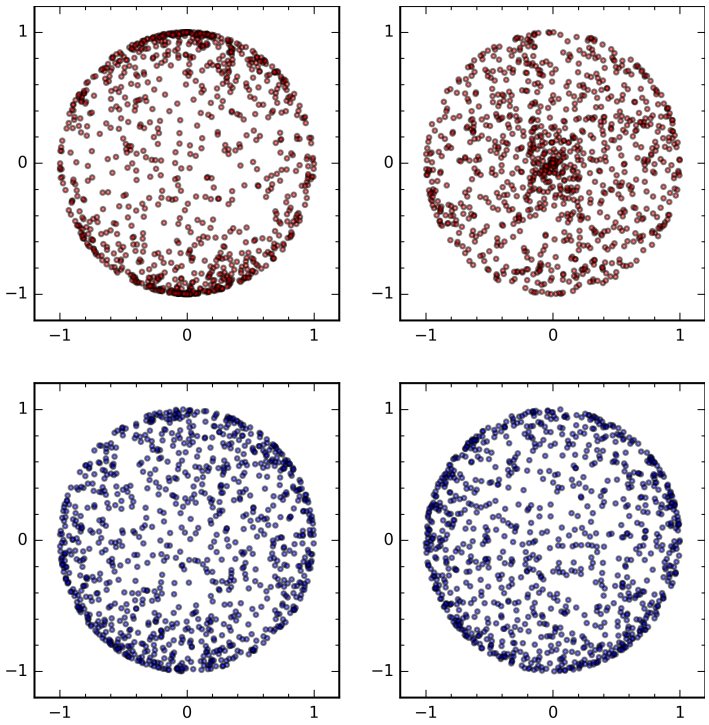


Figure 2.18: Projection on two different planes of a non-uniformly generated direction on the 3-dimensional sphere (top); the same plot is repeated for an uniformly distributed sampling (bottom). In the first case, the points are generated more frequently in correspondence of the poles, while a homogeneous scattering is recovered with the correct sampling.

2.10 Gas-phase cohesive energy on a fcc lattice

In order to check whether a significant energy difference exists for the aggregation of molecular clusters of different compositions into larger supramolecular structures, we decided to model atomistically a bulk system representing a small portion of the vesicles membrane. This system would indeed allow the estimation of the aggregation energy resulting from the cohesion of clusters in a packed lattice. Particularly, we are interested in comparing the energetics of the system by varying the relative concentration of isomers. According to the proposed gelation mechanism, the homochiral system is expected to exhibit a lower interaction energy with respect to the heterochiral one. A confirmation of this fact would support the chiral selective aggregation of clusters, consistently with the experimental evidence.

To obtain such an estimation of the cohesive energy, we devised a system consisting of a *face-centreed-cubic* (fcc) lattice of molecular clusters in periodic boundary conditions. The choice of an fcc lattice was not driven by any experimental finding; however, since it is the denser arrangement of spherical objects, it is expected to be the most adequate representation of a bulk system. The idea behind this model is to be able to gather a broad statistics of the

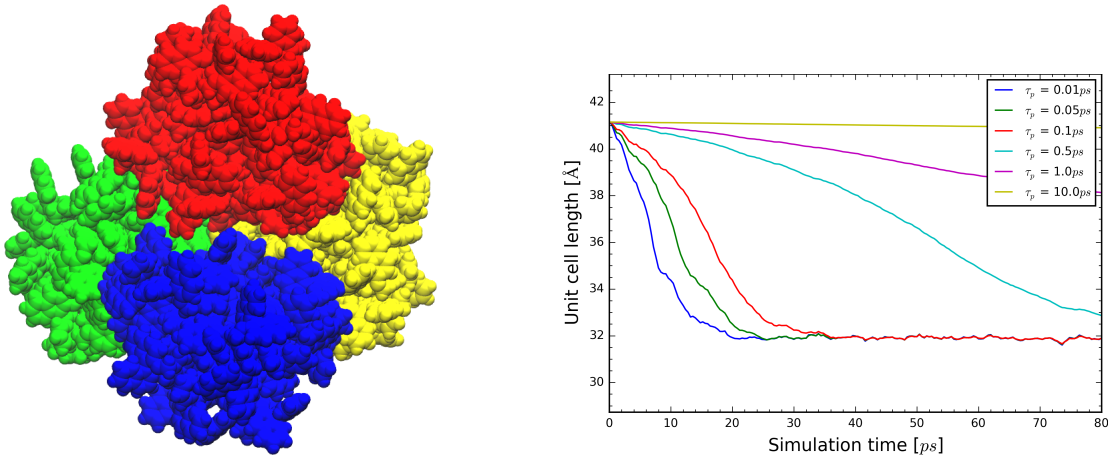


Figure 2.19: Tetrahedral unit cell for an fcc lattice of 10 dimer clusters (left): for enantiopure system, clusters are composed by 10 R,R dimers, while for the racemic one, clusters contain 5 dimers for each chirality (see reference structures on the top of fig. 2.16). The structures are taken from the simulation conducted in sec 2.8. On the right the cell length's decay is plotted as function of the barostat coupling parameters τ_p .

energetics of the system upon relaxation of the lattice in the gas phase. For simplicity, we were not interested in taking into account the contribution of the solvent molecules. Despite the solvent demonstrated to play a role in the gelation mechanism, we decided to focus the attention on the chiral sensitivity of the process, thus explicitly considering the solute only. In fact, the interaction of binaphtol derivatives with the solvent cannot explain alone the complete inhibition of gel formation evidenced in case of racemic compositions. Moreover, the great stability of the gel, especially at high temperatures, suggest that some "locking" mechanism is required to provide enhanced mechanical stability to the aggregates; conversely, this "locking" binding configuration is expected to be missing, or highly defective for racemic composition. In this regard, we remind that since neither experiemntal nor computaitonal results provided robust elements to justify the chiral selective gelation, we have reason to claim that the solvent action is, in case, marginal and limited to a solubility effect.

In order to collect a sufficient statistic out of the bulk model, it is necessary to generate a collection of fcc lattices composed by randomly oriented clusters. Clearly, the relaxation of a single unit cell would not be statistically representative for the lattice cohesive energy. There are two main reasons for this:

1. a unique 10 dimer cluster was not identified for neither the homochiral case, nor for the heterochiral one;

2. the configuration space of the clusters in the fcc lattice has to be explored thoroughly.

Two lattice models were considered: an homochiral lattice, formed by 10 dimer clusters of R,R enantiomers, and a heterochiral one, formed by clusters of mixed composition, containing 5 mirroring enantiomeric dimers for each chirality. Five different structures for each case were selected from the equilibrium simulations described in the previous section, having energy corresponding to the calculated means (cfr. fig. 2.16). The clusters were positioned in a tetrahedral unit cell at a distance of 28 Å, for an initial box cell length of about 41 Å. The generation of different relative orientations within the tetrahedral unit cell was performed with the approach described in section 2.9.1. The starting cell size was intentionally set to a value larger than the one expected after relaxation: this was necessary, since random cluster rotations would inevitably lead to superimpositions of atoms. In fact, the clusters are not perfectly symmetric and a too densely packed initial configuration would result in the occurrence of undesired and unphysical contacts, whose prevention was fundamental to avoid energy peaks affecting the relaxation procedure.

For each initial lattice configuration the relaxation was carried out to further minimise the energy of the system and to eliminate the initial spacing. In order to do so, short molecular dynamic runs of 50 *ps* were performed at 40 K, during which the unit cell was relaxed through the application of an isotropic Berendsen barostat at null reference pressure and Langevin thermostat. The coupling parameter of the barostat τ_p (see sec. 1.2.4) was chosen in order to achieve the relaxation of the lattice in a short time and decrease the overall computational cost. Since the annealing procedure was carried out in the gas phase, the effect of the barostat is limited to the isotropic scaling of position, until convergence of the simulation box size is reached, while little effect is exerted on the clusters.

On the right of fig. 2.19 the length of the unit cell vector is plotted against the simulation time for various values of the barostat's coupling. For τ_p greater than 1 *ps*, the unit cell did not equilibrate within 80 *ps*, chosen here as an upper bound for the duration of the annealing process. By further decreasing the relaxation time, the unit cell converged more quickly, till reaching a cell length of 32 Å. A negligible faster relaxation was instead observed for a value of τ_p lower than 0.1 *ps*. This last value was chosen for this set of simulations since it allowed fast relaxation, while resulting in a compressibility within the range normally used for simulation in explicit water. After the first 50 *ps* the unit cells were subsequently annealed to 0 K and further minimised through a conjugate gradient algorithm, yielding an energy for the lattice E_{fcc}. From the minimised structure, the coordinates of each cluster were extracted and minimised separately in gas phase (no boundary conditions applied in this case) to obtain the energy E_i. The cohesive energy of the lattice, being the energy for a close packed arrangement, can then

#	$\triangle E_{cohe}$ 0% RR [kcal/mol]	$\triangle E_{cohe}$ 50% RR [kcal/mol]
1	-225.5±1.2	-222.3±1.2
2	-231.8±1.6	-217.0±1.3
3	-238.4±1.1	-196.3±1.2
4	-231.6±1.4	-214.3±1.3
5	-233.1±1.2	-237.5±1.1
Mean	-228.5	-217.5

Table 2.6: Cohesive energies for the cluster fcc lattice: the homochiral system displays a slightly lower energy with respect to the heterochiral one.

be easily calculated as

$$\Delta E_{cohe} = E_{fcc} - \sum_{i=1}^{N} E_i, \tag{2.20}$$

where N is the number of clusters forming the fcc lattice, in this case equal to 4. The results are collected in table 2.6 for each system.

In 4 out of five cases, the cohesive energy of a homochiral 10-dimer lattice is lower than that in the heterochiral case. On average, the different composition of the fcc lattices displayed an energy difference of about 11 *kcal/mol* over 5 different cluster geometries. This difference is roughly on the order of magnitude of a H-bond, and is particularly small especially if the number of molecules is considered (80 binaphtols derivatives in total). The standard errors bear information on the fact that each average was performed over a broad statistics of 1600 configurations, thus resulting in a little uncertainty of the means. However, a rather high standard deviation of the plots does not provide enough element to conclude with a sufficient margin of certainty that the reported differences are consequence of the chiral-induced steric effect between atropoisomers. Nonetheless, since these results are consistent with this picture they do not allow one to dismiss it. On the contrary, once again, they point toward the fact that if chirality has a small (yet consistent) effect on average, it might well be that the aggregation leading to gelation is driven by a limited set of binding configurations, where the pure energy contribution is not necessarily dominant.

2.11 Sampling the dimer-dimer interaction energy

The relaxation of clusters and *fcc* lattices, provided some hints on the possible existence of a set of configurations that could rationalise the observed or missing gel formation on a supramolecular

level. However, it was particularly challenging to detect relevant structural features bearing sign of any "locking" interaction: in fact, the visual inspection of the three dimensional structures did not easily allow us to do so, particularly when trying to resolve the inner portion of clusters; moreover, a description of the clusters through an order parameter was demonstrated to be difficult, since the definition of a collective variable, as routinely done for metal clusters, falls short for molecular structures and for a small number of particles i.e. dimers.

To fill the gap left open from the results obtained from the previous studies, we next focused on the interaction between pairs of dimers, in the attempt to determine a set of relevant binding configurations. The supramolecular assembly, indeed, ultimately needs to be explained by the presence of defined directional aggregation patterns. This is, in fact, a necessary condition for the formation of fibres, as highlighted by a number of papers in the literature, most of which, though, did not specifically address the study of supramolecular aggregation at the atomistic level.

2.11.1 The Monte Carlo sampling

The tool for generating random orientations was used within a Monte Carlo sampling algorithm to explore the binding between dimers. Staring from an initial displacement and a random orientation of a pair of dimers (as described in section 2.9.1), the space of the configurations was sampled by performing subsequent random moves. Each move consisted of a roto-translation which is attempted upon checking that the distance between intermolecular atom pair was not smaller that 2 Å, in order to avoid unphysical contacts before even attempting the energy estimation. The translational part of the move was instead characterised by a displacement uniformly distributed in the interval $[0, 1]$. A further check was done on the separation distance between the centres of mass of the two species, in order to limit the maximum separation distance to L=20 Å; this distance, corresponding to the larger dimension of the molecules considered, is small enough to limit the physical volume while sufficiently big to characterise the interaction in full. The rotation is performed by using the transformation matrix in eq. 2.10 for a small value of the angles θ, ϑ, and χ. The acceptance probability was therefore chosen using the standard Boltzmann statistics and it is expressed as follows,

$$P(x_i \rightarrow x_{i+1}) = \begin{cases} 1 & if\ e^{\frac{-(E_{i+1}-E_i)}{kT}} > w_i \\ 0 & otherwise \end{cases}, \tag{2.21}$$

where w_i is a uniform random number generated in the interval $[0, 1]$ and E_i is simply the energy of the (attempted) *i-th* configuration. For each rejection, a new move is attempted until

a new configuration is found. If no configurations are accepted after $M = 500$ attempted move, another starting configuration is generated and another Monte Carlo chain is performed over the system. The energy of each accepted configuration is stored, allowing for the calculation of ensemble average energies; moreover, the final configurations for each Monte Carlo chain was further annealed and minimised at 0 K to find the local minima of the PES.

This algorithm was implemented within a Python script, which was used to generate the configuration for each move and call an external MD engine (AMBER) for the energy evaluation.

2.11.2 Binding configurations of dimer A

In this section, the results of the calculations performed for pairs of dimers (type A) are presented. The main purpose of these calculations was to find out whether the reason behind the missing gel formation could be evidenced by the analysis of the binding of pairs of dimers. In this framework, particular attention was devoted to the comparison of the configurations resulting from the interaction between two homochiral dimers against two enantiomeric ones, in order to establish whether (i) the minimum value of the binding energy is affected in any way by the chirality and/or (ii) the binding mode corresponding to those minima is associated to distinctive supramolecular architectures from which to infer possible directional patterns. We considered the following cases:

- the interaction between two homochiral dimers;
- the interaction between one R,R-R,R dimer and a S,S-S,S dimer;
- the interaction of two homochiral methylated binaphtol derivatives.

While the first two cases are meant to elucidate the aforementioned chiral factor, the third is introduced to study the effect of the methylation. As reported in section 2.7, the presence of the methyl group is indeed sufficient to inhibit the gel formation completely. This is quite a remarkable effect that provided another case useful to rationalise the mechanism leading to gel formation.

A Monte Carlo approach was chosen to explore the orientational space at 300 K, as described in section 2.11.1. More than three thousand Monte Carlo chains were performed on each of the three systems. The evaluation of the energies for each move were implemented on the basis of the same parametrisation; the solvation energy was modelled by a generalised Born implicit solvation model presented in sec. 1.3.2, consistently with the parametrisation used in [81]; an implicit solvation scheme in place of an explicit one was preferred since it allowed to reduce the

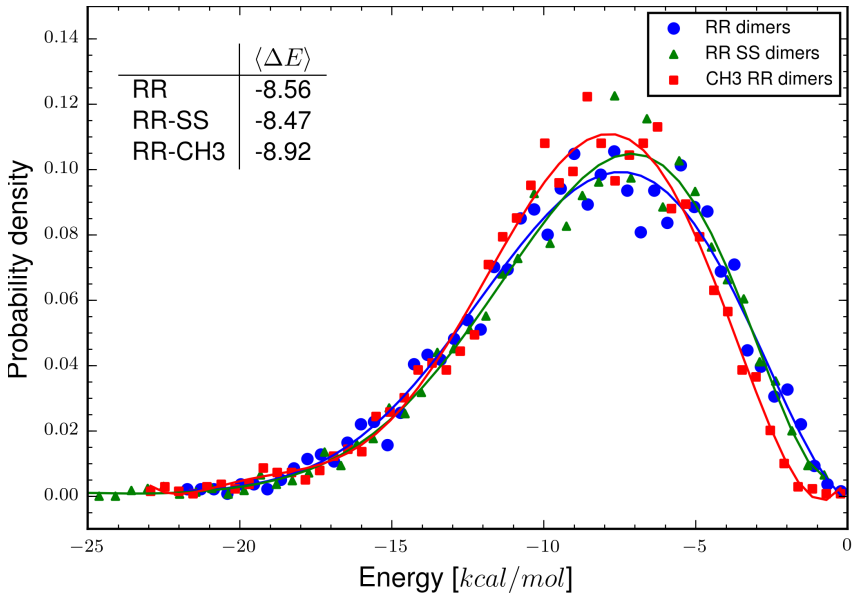


Figure 2.20: Energy probability densities calculated from a Monte Carlo sampling at 300 K for the three systems.

complexity of the attempted Monte Carlo moves by essentially excluding the solvent molecules' degrees of freedom; since methanol was chosen as a reference solvent, a dielectric constant of 32.7 was used. The binding energy was then calculated by subtracting from the energy values obtained the energy of each dimer calculated separately at the same conditions (see section 2.4.1).

The results of the Monte Carlo simulation at 300 K are reported in fig. 2.20. The three probability densities are represented by a polynomial fit of energy distributions gathered by sampling the configurational space. The plots do not display significant differences: the blue, green and red curve refer to the homochiral, heterochiral and methylated pairs, respectively. The peaks of the distribution are located around the same energy value, corresponding to about -7.5 *kcal/mol*. For the methylated dimer, a slight offset can be observed, corresponding to roughly 0.50 $kcal/mol$, a difference that, however, was not considered significant. The structural diversity of the dimers does not allow, therefore, to appreciate significant difference in the interaction of dimer, as far as averages are concerned. The average binding energies, calculated through the polynomial fit of the scatter plots, are reported in the inset of fig. 2.20, further confirming the equivalence of the energy distributions. The reported average corresponds to about -8.5 $kcal/mol$ for RR and RR-SS pairs, while, for the methylated dimers, a binding energy lower than the former value by 0.4 $kcal/mol$ was calculated.

The average binding energies displayed in the inset of fig. 2.20 with those reported in fig.

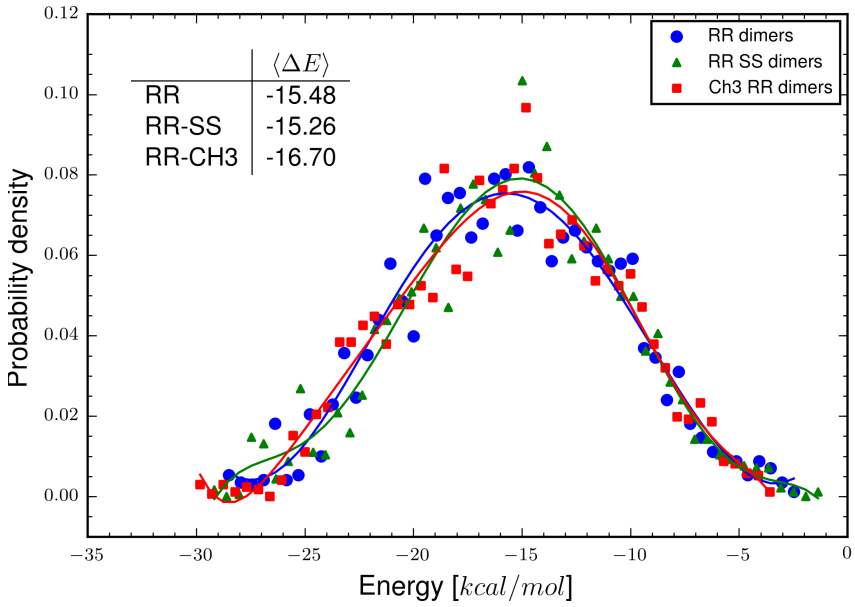


Figure 2.21: Energy probability densities calculated from further minimising the minima found with the Monte Carlo mothod.

2.5 would suggest that the interaction of dimers is not favoured over the direct aggregation of single binaphtols, since the values reported in the first case are lower by about 2 $kcal/mol$. A direct comparison, however, is likely biased by the different sampling method used, and the different solvation model (implicit vs explicit). The reason for the adoption of a distinct strategy method for the sampling of dimer-dimer interactions is justified by the difficulty of sampling homogeneously the orientations for a adequate simulation time with MD, while assuring the cohesion of the dimer structures.

The final configuration obtained from each Monte Carlo chain was annealed by means of short dumped dynamics runs of 30 ps followed by conjugate gradient energy minimisation. The resulting binding energies are representative of all the local minima in the configurational space at 0 K. The energy densities are reported in fig. 2.21, analogously to what was done earlier. Also in this case, no significant differences were in evidence, given that the fitted probability densities do not show any clear shifts. Furthermore, the averages calculated from the respective densities, reported in the inset of 2.21, confirm that no difference can be detected at this level. For both the heterochiral and the homochiral pairs, binding energies of -15.5 and -15.3 $kcal/mol$ were respectively obtained. Again, for the methylated dimer a slightly higher binding energy was found, about 1.5 $kcal/mol$ higher than the two previous cases. This is consistent with the fact that, the introduction of a methyl group results in additional attractive interactions (missing for the other two cases) that marginally increase the stability of the methylated dimer pair.

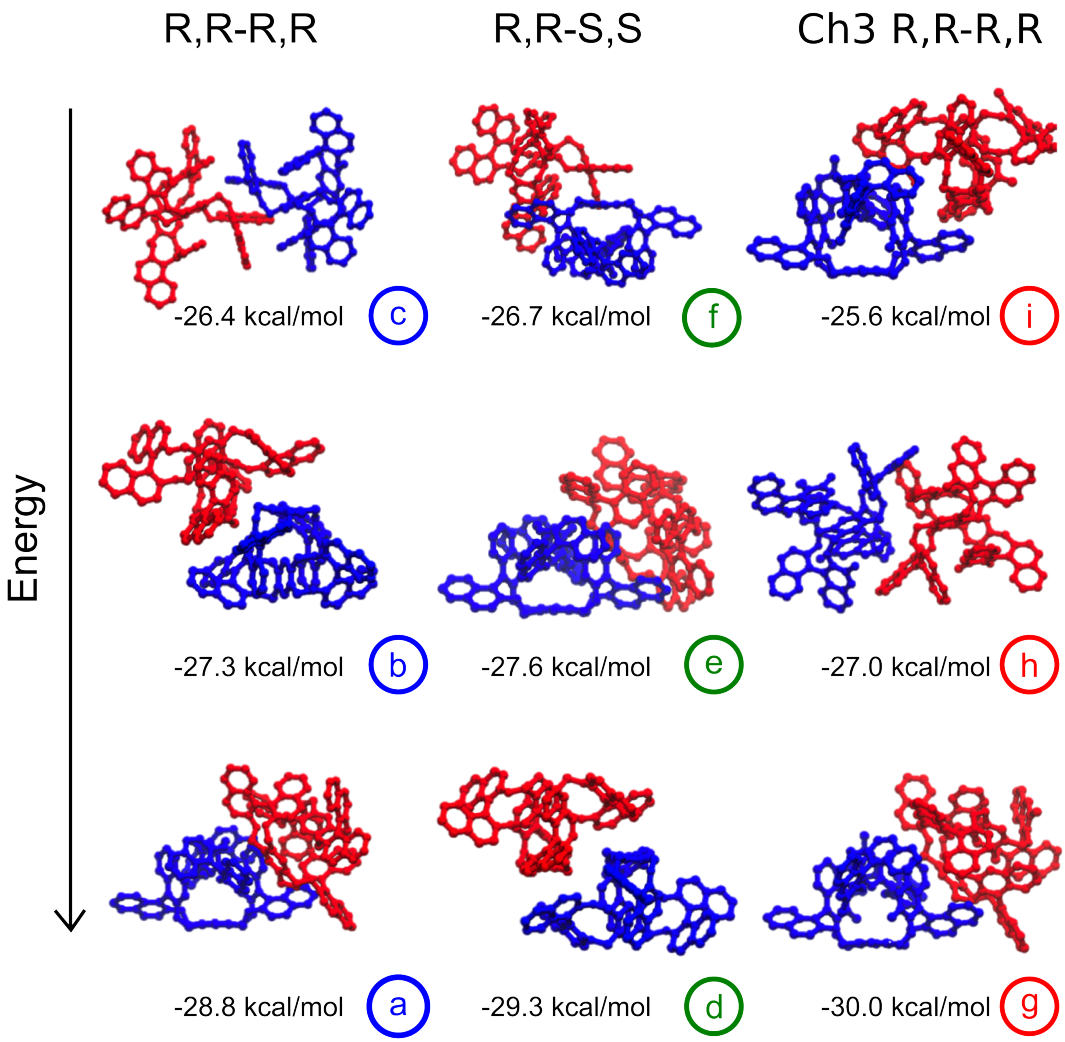


Figure 2.22: Configurations of minimum binding energy for three pairs of dimers: pair of R,R dimers in first column (structures a, b, c), R,R and S,S dimer pairs in middle column (structures d, e, f), Ch3 R,R-R,R dimer pair in third columns (structures g, h, i). The configurations are reported in order of decreasing energy, from the top to the bottom; hydrogens are not shown for clarity.

The analysis of these large sets of dimer pair configurations and the relative binding energy distributions was not proven to be conclusive. Part of the problem, as outlined earlier, implicitly relies on the different phase space sampling of Monte Carlo approaches when compared to classical MD. However, even if no significant difference is found in terms of ensemble average, the complexity of the supramolecular aggregation demands further analysis of the structures obtained upon minimisation. Thus, the last 100 configurations were then aligned according to a common orientation in order to try to detect potential analogies. For each system, the most meaningful configurations were then selected and reported in fig. 2.22, in order of decreasing binding energy from the bottom to the top.

We have identified three main configurations for the RR dimer pair. Starting from the bottom of the left column of fig. 2.22, the configuration a was found to be the most stable, having a binding energy equal to -28.8 $kcal/mol$. The first dimer binds to a naphtyl moiety belonging to the folded conformer of the second one. Two characteristic interactions are present: a first $\pi-\pi$ stacking between the naphtol groups of two folded molecules; a second $\pi-\pi$ stacking between the central phenyl moiety of one dimer and the naphtol group of the extended configuration on the opposite dimer. In this binding configuration, the two dimers are nicely bound with a neat geometrical alignment of the aromatic rings, resulting in a stable, low energy configuration. Configuration b is the second most favourable one; in this case, the two dimers are antiparallel and they interact through a double symmetric $\pi-\pi$ stacking between the extended conformer and the phenyl ring of the folded one. Despite its symmetry, this binding is characterised by a less precise alignment of the aromatic rings, resulting in a smaller binding energy, equal to -27.3 $kcal/mol$.

The third configuration, labelled c, is very interesting. Here, the dimers are positioned at a higher distance, resulting in a smaller binding energy; however, they display an interesting binding geometry: the extended conformers are bound through offset parallel $\pi-\pi$ stacking of two naphtyl moieties, while two $ch-\pi$ bonds are simultaneously formed between the remaining two naphtol groups and the hydrogen of the folded conformer pointing towards them. The resulting alignment allows a highly symmetric positioning of the dimers with a clear directionality and possibility of further extending the binding pattern in a reperitive manner.

For the heterochiral dimer pair, three configurations are reported in fig. 2.22. Two out of three present a remarkable similarity with two configurations found for the homochiral dimers. The configuration d is, indeed, analogous to the configuration b reported earlier: the dimers are again oriented in an antiparallel way but, differently to the homochiral case, this is the most stable configuration, having a binding energy of -29.3 $kcal/mol$. Conversely, the configuration e, which presents the same binding motif of configuration a, has a higher binding energy than

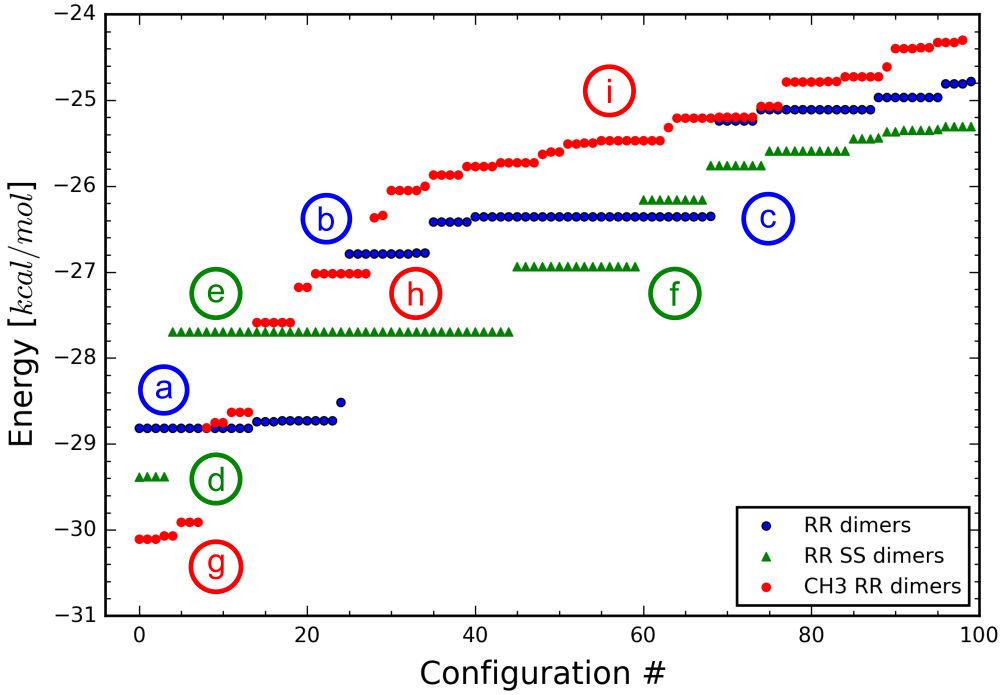


Figure 2.23: Last 100 configurations of minimum binding energy for the dimer pairs. Each configuration illustrated in fig. 2.22 is labelled here by the corresponding letter and colour, showing the different energy level and their population.

the antiparallel alignment, equal to -27.6 $kcal/mol$. The last configuration, f, has no homochiral analogue. It originates from the formation of two $\pi - \pi$ stackings between the central phenyl groups of folded conformers and the naphtyl groups of the extended conformer of opposite dimers. This binding geometry have a binding energy of -26.7 $kcal/mol$.

Three configurations are reported also for the methylated dimer. The binding geometry g presents an orientation analogous to a and e of the homochiral and heterochiral dimer pair, respectively. The cohesive energy is even higher in this case than that calculated for the corresponding homochiral dimers (-30 $kcal/mol$), possibly due to the additional contribution arising from the methyl groups. The remaining two structures h and i are representative of a set of highly populated binding modes, whose fine structure is not displaying any characteristic aggregation pattern. The binding energies is nonetheless comparable with the values already reported, being equal to -27.0 and -25.6 $kcal/mol$, respectively.

The minimum energy configurations reported here for pairs of dimers allows one to make some observations. First of all, there is a recurrent configuration that is found for all systems, corresponding to the binding a, e and g. The binding energy of this configuration is the lowest found for the homochiral cases, both for the native binaphtol and the methylated one, while for the

heterochiral dimers it is the second most energetically stable one. A different energy minimum was found for dimers of opposite chirality, corresponding to the antiparallel geometry displayed in *d*. In particular, we observed that the opposite chirality has the effect of reversing the order of the binding in *a* and *b*, resulting in the *d* and *e* configurations.

Bearing in mind that the gelation occurs neither for heterochiral system, nor for the methylated binaphtol derivatives, the configuration shown in fig. 2.22 c was identified as a candidate building block of the gel network. In fact, this particular structure is only found when considering homochiral dimers and its symmetric binding motif displays an ordered set of $\pi - \pi$ stacking interactions supporting the stability of the gel to high temperatures. Despite not being the most energetic one, it satisfies the condition for causing the formation of larger aggregates.

The methylation of native homochiral binaphtols has the effect of altering the dimers' geometry when compared with the non-methylated cases. Both the antiparallel configuration *b* and *c* were indeed not reported (among the most favourable 100 configurations): various binding geometries within the same range of energies were instead found, none of which displayed comparable binding features, nor the possibility of originating repetitive patterns. An hypothesis for this can be drawn: the modification of the binding sites achieved by the insertion of methyl substituents alters the geometry of binding sites and consequently the stability of the minima observed for the non-methylated molecules. It is worth noticing that all the structures reported in fig. 2.22 represent just one possible arrangement of the respective dimer, while each of them has two equivalent binding sites, yielding a total of four equivalent binding orientations. When dealing with methylated binaphtols, the four equivalent binding geometries are no longer degenerate. This does not only affect the binding for two dimers, but also the possibility of forming repetitive arrangements, once a third dimer is added.

The binding energies of 100 lower energy dimer pairs are reported in fig. 2.23, where each letter identifies one of the configurations reported in fig .2.22, and it is displayed in the graph for increasing binding energy. The data shows that the sampled binding energies appear to be segregated within discrete energy levels (this "quantisation" is particularly evident in the homochiral case); the population of each level provides a measure of the relative stability of each configuration. The binding energies of homochiral dimer pairs are displayed in blue: essentially, the three configuration found in *a*, *b*, *c* correspond to as many distinct energy levels. In particular, the binding *a* and *c* are associated to the most frequently observed, confirming the stability of the minima *a* and, particularly, supporting the hypothesis that *c* could be a potential building block for the assembly of the gel phase. For the heterochiral dimers, the antiparallel configuration is the lower energy one, but it was observed just in four cases. On the other hand, configuration *e*, whose binding energy is lower by $\sim$1.7 $kcal/mol$, is by far the most frequently

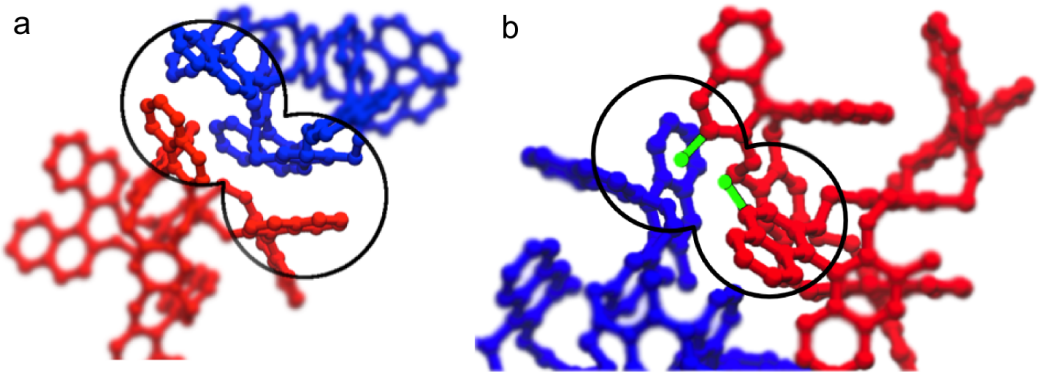


Figure 2.24: Image of the binding site for configuration e of the heterochiral pair (a) and configuration g of the methylated moiety (b); methyl groups in green refers to the unfavourable symmetric configuration (hydrogens are not shown for clarity).

counted. For the methylated dimers, instead, the situation is rather different. Indeed, it is not possible to observe an analogous separation of the energy levels as for the other two cases. On the contrary, more levels are present accounting for the appearance of different configurations that could not be easily grouped according to their orientations, as briefly mentioned earlier. This is consistent with the aforementioned loss of symmetry in the binding interactions caused by the presence of the methyl groups.

We have observed how different chirality or small structural changes of the native molecules produced measurable differences in the binding energy and geometry of binaphtol's dimer units. This fact is particularly relevant since $ch-\pi$ and $\pi-\pi$ stacking interactions are highly directional and, therefore, highly sensitive to the relative orientation of the aromatic rings. This is highlighted in fig. 2.24, where two examples of the binding site of configuration e and g are displayed. In the homochiral case, this binding configuration involves the offset parallel stacking of two naphtyl moieties and of a naphtyl with a phenyl centre (conf. a of fig. 2.22). In fig. 2.24 (a) a screen shot of configuration e for the heterochiral dimers is reported. At difference with the homochiral case, the contemporary alignment of both the aromatic groups cannot be achieved. As a result, the binding energy is lowered by 1.2 $kcal/mol$. In fig. 2.24 (b) a close view of configuration g is also reported. In this case, the binding geometry is only marginally modified since the methyl groups do not affect the geometry of the binding site. However, we observe that their position is such that the degeneracy of the symmetric configurations is lost. This situation is illustrated by the two fictitious methyl groups rendered in green in fig. 2.24, which aim at representing the symmetric binding geometry: steric hindrance is expected to obstruct the formation of $\pi-\pi$ stacking, thus affecting considerably the binding geometry that, in fact, was not observed during the simulations. This provides an important argument for the

asymmetric binding feature introduced by methylation, which is believed to prevent gelation.

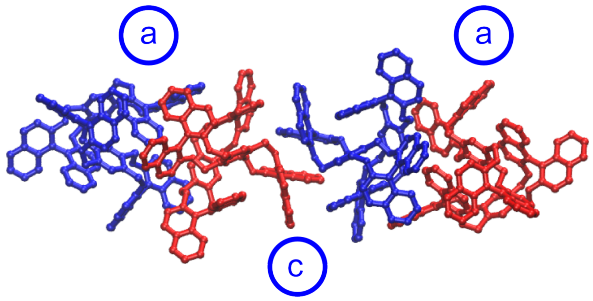


Figure 2.25: Two pairs of dimers on the binding configuration a of fig. 2.22 were found to form a stable interaction through configuration c (hydrogens are not shown for clarity).

The minima found for dimer pairs reported above were then used as input structures for the same Monte Carlo algorithm. The aim was to verifying whether a directional and repetitive aggregation motif could be detected on the base of the interaction subsequently highlighted. Importantly, when the possible orientation of two RR dimers in configuration a of fig. 2.22 were sampled, it was found that the dimers could bind via a c type interaction (see fig. 2.22). This is an additional evidence supporting the relevance of binding c for the aggregation of dimers, particularly since it clearly displays the sort of directional and repetitive aggregation motif - of the type $-[c-a]_n-$ which can be potentially associated to the formation of "gel fibrils". In fact, this aggregation pattern was not detected when studying the coalescence of heterochiral or methylated dimers. Further checks, in fact, showed that such aggregation is partially hindered for heterochiral dimers for steric reason and does not lead to the same neat alignment of molecules, thus yielding a significanlty less stable structure.

2.12 Ising-like model for aggregation

The results of the previous section provided evidence that a set of specific interactions between homochiral dimers could rationalise a possible scenario for the directional aggregation of dimers in a higher order, directional structure through a $-[c-a]_n-$ polymerisation mechanism. Importantly, this mechanism would not be possible during the aggregation of either heterochiral or methylated binaphtols, since they cannot establish type c binding. This is in full agreement with the experimental evidence for which only the homochiral molecule could form a gel phase through the formation of fibrils. To further test this hypothesis, a simple lattice model was devised. The system consists of a 50x50 lattice where each site identifies the position of a dimer, being either a homochiral or a heterochiral (methylated) one. The particle type occupying each

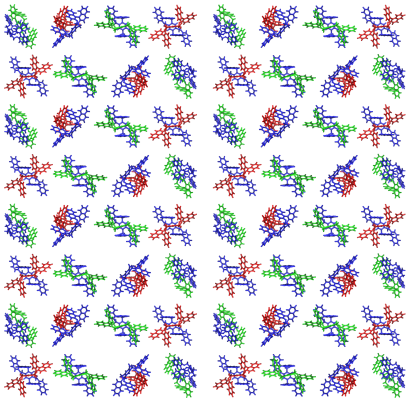
site is represented by an integer n, assuming a value of 0 or 1, for the homochiral and heterochiral (methylated) species, respectively; the ratio of zeros and ones corresponds to the fraction of homochiral/heterochiral dimers, labelled as X_{RR}. The dimers at each site can have distinct orientations: the coupling between neighbouring molecules could be in principle calculated by associating the binding energies obtained with our MC scheme to some relative binding directions. Given the high dimensionality of the PES, machine learning methods (decision tree based regression models) could allow for improved computational efficiency, in particular by enabling the full definition of a lattice site with a set of two vectors. Moreover, this approach would allow us to overcome the precise geometrical description based on a fully atomistic force field, in favour of a simplified but still exhaustive vectorial representation. To some extent this fitting of the pair interaction energy is reminiscent of the Gay-Berne potential for asymmetric colloidal particles [82]. However, the intrinsic nonlinear dependence of the energy upon rotation and the discontinuity caused by molecular topology led to an inaccurate energy interpolation, whose average error exceeded 15 % (about 2.5 kcal/mol).

To circumvent this issue, a simpler approach for the representation of the possible orientation was preferred. The configuration of the dimers was represented by an integer, k, assigned to each lattice site. These integer numbers were chosen within a set of values representing the orthogonal direction in space ($\leftarrow, \rightarrow, \downarrow, \uparrow,$). The total energy of the system was that determined as the the sum over the nearest neighbors of the pair interaction energy $E(i,j)$ defined as follows:

$$E(i,j) = \begin{cases} J & if\ |\Delta k| = 1 \\ 0 & otherwise \end{cases} \tag{2.22}$$

where J is an energy parameter representing the interaction contribution: for simplicity J was set to kT, for a reference temperature of 300 K, thus corresponding to roughly 0.597 *kcal/mol*. While not corresponding to any of the binding energies previously calculated, this coupling factor accounts for energy differences of the dispersive attraction arising between different orientations, within the range to those found earlier. The sum is expanded over the whole lattice using periodic boundary conditions; since the system is neutral and the dimension of molecules is large enough to exclude the influence of non-local vdW interactions, no long-range term was considered necessary which, on the contrary, were shown to greatly influence aggregation in presence of dipoles and charges [83]. Each particle of configuration k could thus favourably interact with the neighbouring ones, having configurations with label $k-1$ and $k+1$. To mimic the effect of a different chirality, the set of indices representing the possible configurations of the dimers was chosen from different non-overlapping ranges of integers. This was done with the

idea of representing heterochiral dimers as having a set of binding orientations that could, for the Hamiltonian reported in eq. 2.22, hinder the formation of a sequence of repetitive interactions $-[c-a]_n-$ by reducing the number of favourable combinations between heterochiral (or methylated) dimers. As experimentally observed, indeed, homochiral dimers are expected to arrange themselves in stable membrane-like structures that are sensitive to enantiomeric excess in a way that could not be rationalised by the analysis of the binding energies of pairs of dimers. This model presents similarities with those used for the study of critical phenomena such as glass transition and percolation theory [84]. The model does not include solvation effects, as the solvent nature was found to play a minor role in the gel formation.



(1,4)	(0,5)	(1,2)	(0,7)	(1,1)	(0,6)	(1,1)	(0,8)
(0,5)	(1,3)	(0,5)	(1,3)	(0,6)	(1,1)	(0,7)	(1,2)
(1,2)	(0,5)	(1,3)	(0,8)	(1,4)	(0,6)	(1,2)	(0,8)
(0,7)	(1,2)	(0,5)	(1,2)	(0,6)	(1,1)	(0,7)	(1,4)
(1,4)	(0,5)	(1,2)	(0,7)	(1,2)	(0,6)	(1,3)	(0,8)
(0,5)	(1,1)	(0,6)	(1,3)	(0,6)	(1,2)	(0,8)	(1,2)
(1,1)	(0,8)	(1,2)	(0,7)	(1,1)	(0,6)	(1,3)	(0,5)
(0,7)	(1,1)	(0,7)	(1,3)	(0,5)	(1,2)	(0,7)	(1,1)

Figure 2.26: Schematic representation of the lattice model for the gelation mechanism: the arrangements of dimers in a square lattice (right) is represented by a matrix where each particle is described by two integer number identifying the molecular type and the orientation.

The physical system and its respective index matrix are represented schematically in fig. 2.26: each site is occupied by either a homochiral or heterochiral dimer with an orientation expressed in the matrix with the couple (i, k). From an initial randomly generated configuration, the statistics of the system is sampled through a Monte Carlo algorithm. At each step one of the following moves is attempted:

- assignment of a new k value within the correspondent interval for a randomly selected particle; this represents an attempted rotation;
- exchange of position between two neighbouring particles.

For both the moves, only the position or the integer k is modified while the fraction X_{RR} is kept constant. The acceptance probability (see section 2.11.1) is chosen as to guarantee the sampling of the Boltzmann distribution (see eq. 2.21). A set of simulations were conducted for increasing enantiomeric fraction X_{RR} from 0 to 0.5, corresponding to the homochiral and

heterochiral systems, respectively. The simulations were carried out for each system in a range of temperatures between 210 and 450 K. The results of each simulation were averaged over 5 independent sets of runs for each temperature and composition (i.e. same X_{RR}): this was done to ensure that the results are not affected by the choice of the initial configuration. Each Monte Carlo run consisted of an equilibration stage, followed by a run of 10^7 steps, where the latter was used for the analysis. An equivalent lattice model was also performed on a hexagonal lattice, for comparison (not reported here for no significantly different results were in evidence).

2.12.1 Results

The simple model represented herein, provided interesting results that allow us to outline in a simple way the behaviour of the real system. Fig. 2.27 shows the average energies calculated as a function of the composition and for increasing temperatures. The energies are expressed in reduced units and have been divided by the number of lattice sites (N=2500). In the low temperature range, the per particle energies are near the global minimum for the chosen Hamiltonian, since the system is allowed to sample only those states that are close to the minimum. At higher temperatures, the average energies increase progressively, since the statistics favour more diffusive states of the systems.

The most distinctive feature of the plots is the energy dependence on the systems' composition. In fact, at all temperatures an increasing value of X_{RR} corresponds to a higher average energy. At the lowest temperatures tested, equal to 0.7 (210 K), a variation of 10% was observed between an enantiopure (X_{RR}=0) and a racemic fraction (X_{RR}=0.5); at higher temperatures, this difference is further increased: for a temperature equal to 0.8 (240 K), the energy increases with X_{RR} by nearly 20%, while a variation of 50% is reached for higher temperatures. This evidence is rather remarkable, as it allows a better understanding of the role of chirality within the gel formation mechanism. In fact, this simple aggregation model provides robust theoretical evidence supporting the superior gelation activity of enantiopure solutions.

In fig. 2.28 the lattices for two different values of temperature and composition are rendered according to the indices k and n, being the orientation and molecular type, respectively. For X_{RR}=0.05 and T=0.7 (bottom-left panel), the arrangement of the particles, represented via k, reached an almost-continuous checkered configuration, which we identify as being the proposed gelation pattern induced by the $-[c-a]_n-$ repetitive binding. The effect of a small amount of particles mimicking chiral asymmetry is limited in this case. On the contrary, for X_{RR}=0.5 (top-left panel), a different situation is found: particles segregate into clusters of prevailing chiral composition, as can be evidenced by the comparison with the plots on the right hand side

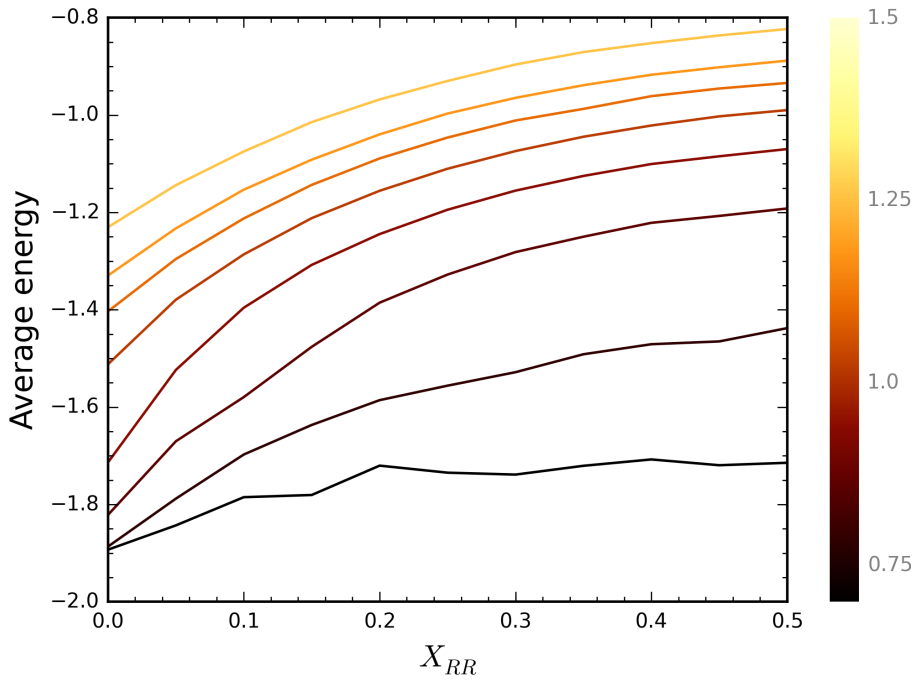


Figure 2.27: Average energy of the lattice model as a function of the fraction X_{RR}. Each line corresponds to a different temperature, labelled in reduce units from 0.70 to 1.50. The energy drops for lower fractions X_{RR}, meaning that the aggregation is more stable both at low and high temperature for an homochiral system.

of fig. 2.28 for integers i and k; this arrangement results in the observable energy difference (see black line of fig. 2.27) of about 0.2 *kcal/mol* per lattice site on average, corresponding to a less favourable aggregation energy of about 10%. As the temperature increases, a disordered arrangement of particles prevails: as the results show, no aggregation motives can be discerned and for X_{RR}=0.5, particle are sparsely distributed. This can be rationalised by the fact that the temperature affects the free energies by weighting the entropic contribution in such a way that less likely configurations (ordered) are penalised if the coupling energy is not sufficient to "quench" some ordering.

To further characterise this model, the specific heat was calculated to monitor the dependence of energy oscillations on the enantiomeric composition X_{RR}. For the present system, the specific heat can be easily calculated using the formula:

$$c_v = \left(\frac{\partial E}{\partial T}\right)_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{NkT^2}, \tag{2.23}$$

where E is the total energy, N is the number of lattice sites and T is the temperature. The specific heat as a function of the temperature can be used to verify the existence of a phase

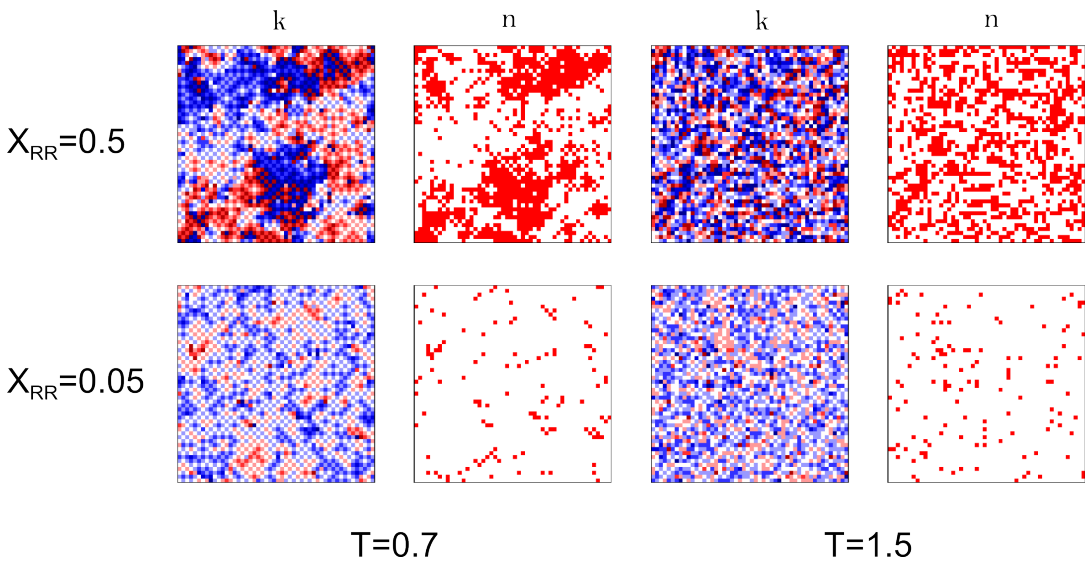


Figure 2.28: Equilibrated configurations for the lattice model for two fractions X_{RR} and two temperatures; the k (=1,...,8) and n (0 or 1) indexes are used to render in different color the orientation and types of particle occupying each lattice site. Lower temperatures (left panels) favour ordered configuration, while a higher temperature (right panels) corresponds to disordered systems, for which the aggregation motives are lost.

transition, that would appear as a discontinuity/peak of the $c_v - T$ plot. In particular, the position of a peak identifies the critical temperature at which the transition is expected to take place.

The values of c_v for each composition are reported in the heat map of fig. 2.29. The bright yellow areas correspond to larger value of c_v, whereas the portion displayed in red represents lower values. In this respect, the heat map suggests a phase transition does exist and is localised in correspondence with the peak defined by the yellow areas. Importantly, the phase transition temperature is found to be dependent on the enantiomeric fraction X_{RR}. For a enantiopure system, in fact, the critical temperature is found for a reduced temperature equal to 1, about 300 K; however, for increasing fractions of X_{RR}, the critical temperature is lowered significantly: for $X_{RR}>$ 0.25, it is in fact found in correspondence with the interval 0.8-0.9, approximately 240-270 K. This fact demonstrates that, for the present lattice model, the stability of an ordered phase (below the critical temperature) is significantly reduced over a range of compositions other than those close to the enantiopure case.

These results complement the average energy dependence previously reported and they further suggest that the presence of a small fraction of molecules not allowing for the continuation of a directional aggregation pattern affects the stability of aggregation itself, here modelled by a simple two-dimensional lattice. In the specific case, the critical temperature shift represents a

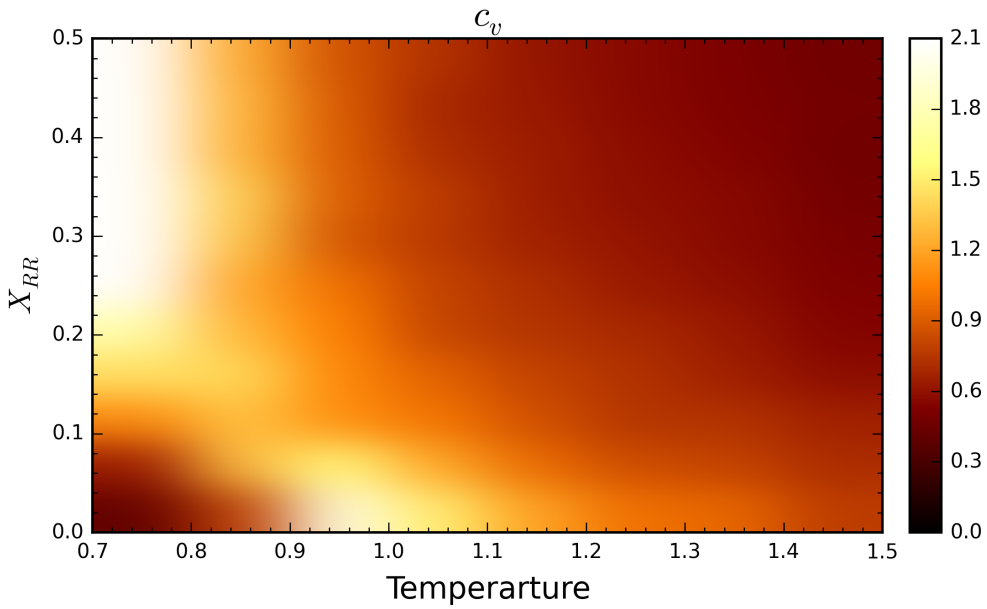


Figure 2.29: Specific heat map as a function of the reduced temperature of the system and the fraction X_{RR}. The brighter areas coloured in yellow identify the peaks of c_v identifying the position of the transition temperature. Importantly, the critical temperature shifts towards lower values for increasing X_{RR}. This indicate that the system is characterised by higher stability when enantiopure.

situation in which the enantiopurity of the composition, i.e. $X_{RR} \sim 0$, enhance the stability of an ordered phase, as displayed in the bottom left panel of fig. 2.28, over a larger interval of temperatures. For composition closer to the racemic one, instead, two separate phases are formed, each characterised by a prevalent composition of one of the two particle types (cfr. top left panel of fig. 2.28); this phase segregation does not persist for temperatures comparable with the enantiopure case, and a disordered phase is formed at just 300 K.

2.13 Summary

The work presented in this chapter was aimed at characterising the interactions between chiral poly-aromatic gelators through both MD and Monte Carlo models, which were used to rationalise the experimental findings. Particular attention was devoted to provide an explanation for the sensitivity of gelation to the molecular chirality. We showed that the gelators can assume two main conformations which have similar energies, thus expected to be simultaneously present in the solution at equilibrium; moreover, the conformational transition between the two conformations was shown to correspond to a value between 8 and 16 *kcal/mol*, suggesting that the transition should not to be considered a rare event. Thus, it is not expected to be directly

connected with the gelation mechanism as some authors have suggested in various occasions for similar systems [60, 58, 59].

The interactions formed between various pairs of molecules were subsequently studied: the molecules preferentially interact via the formation of $\pi - \pi$ stacking interactions between the naphtyl groups and the central benzene ring. The lowest average binding interaction energy was calculated for pairs of homochiral trans- conformers: importantly, the energy for heterochiral trans- isomer (extended) pair was much lower. This finding clearly revealed how chirality does affect the pair interaction energy of these molecules. Moreover, the most favourable binding configuration for the homochiral atropoisomers was not reported for heterochiral ones and, importantly, displayed the possibility of forming a repetitive aggregation pattern into a helix: the existence of this aggregate, however, was not supported by further simulations and was, therefore, discarded.

Bulk simulations of the gelator molecules in methanol did not provide to evidence for significant aggregation patterns beyond those one already determined. However, the analysis of the trajectories revealed the existence of two stable dimers, one formed between two homochiral cis-isomers, and a second formed between a cis- and a trans- isomer. Both dimers are characterised by distinctive directional $\pi - \pi$ and $ch - \pi$ interactions. Moreover, the dimers showed a significantly decreased energy for the respective heterochiral structures. The potential importance of the dimers in the aggregation mechanism was further examined. Firstly, the energetics of clusters of 10 dimers were evaluated through MD, highlighting that their energy does not depend strongly on the chiral composition. On the other hand, the aggregation of 10 dimer clusters modelled via the relaxation of an fcc lattice of clusters displayed a cohesive energy (thoroughly averaged over the orientation space) that indicates some preference in the aggregations of homochiral clusters instead of racemic ones.

The interaction arising between pair of dimers was studied with an *ad hoc* Monte Carlo algorithm exploiting an uniform sampling of orientations. Again no average energy dependence was found for dimers of different type. However, the analysis of the most favourable interactions suggested the existence of a possible sequence of alternating binding configurations leading to a repetitive aggregation motif. In particular, one of the configurations was not observed between mirroring dimers and between the dimers formed by the methylated binaphtols derivatives (for which gelation was not observed in experiments). In this respect, a lattice Monte Carlo model aiming at modelling the directional aggregation, provided insight into the energy dependence resulting from the chiral composition. This model showed how the average energy of the system is expected to increase significantly with the enantiomeric fraction, with the equilibrium state varying from a ordered arrangements of binaphtols to a phase separation in domains of prevailing

chirality. Consistently, the phase transition temperature was observed to decrease for a racemic composition, thus favouring a less ordered, less stable state at lower temperatures with respect to an enantiomeric pure system, thus further supporting the proposed mechanism for gelation.

The work here presented provides some important elements towards the comprehension, on a molecular level, of the aggregation patterns leading to the puzzling phase behaviour of binaphtols solutions and it represents a unique contribution to rationalise a complex and important phenomenon such as gelation.

Chapter 3

$\pi - \pi$ stacking for polyaromatic substrates

This chapter gathers the work that has been carried out over the course of the last three years in the framework of two related projects in collaboration with BP. The main idea behind this research was to combine DFT and classical potentials to provide insight to the development of hydrocarbon mixtures. In this framework, the aid of reliable atomistic models proved to be a useful tool for highlighting new directions to improve fuels and lubricants' research.

Here, we present the research conducted for the characterisation of $\pi - \pi$ interactions between small poly-aromatic hydrocarbons (PAHs), which are used for the functionalisation of molecules able to interact with heterogeneous aromatic substrates. A recently developed correction for dispersion forces was used in DFT calculations to determine the binding energies of a series of PAH dimers; the results were compared with the values obtained for the relaxation of the same series of molecules onto a graphene sheet, revealing how smaller PAHs bind more strongly to an aromatic surface than larger ones, once the binding energy per carbon atoms is compared. This suggested a viable strategy for the design of molecules conceived to stick onto aromatic substrates by exploiting multiple small substituents in place of fewer larger ones. An aliphatic dendritic molecule was then proposed and modelled using a correction of the *ff* for the aromatic interactions, based on the fitting of DFT energy profiles. We tested the ability of this molecule to stick to an aromatic surface in presence of solvent molecules and for a variable degree of aromaticity of the surface. Remarkably, we observed that the proposed structure was able to anchor efficiently to the model surfaces, supporting the multivalent effect of this dendritic molecule and the efficacy of small terminal moieties. Importantly, we show that the layering of

hydrocarbon molecules onto the surface plays an important role in the sticking mechanism and it is expected to influence the kinetics of the adsorption process.

3.1 Introduction

The research activity conducted in collaboration with BP was structured into two consecutive projects. The first one was entitled *Spring molecules* and focused on the possibility of identifying a mechanism and a molecular structure capable of storing mechanical energy to be made available upon a rapid change of temperature or pressure, such as those experienced by fuels during the injection process in the combustion chamber. In this respect, the desired additional release of energy was not meant as a way of directly increasing the amount generated during oxidation processes, for which the exothermic contribution of oxidative reactions are by far leading; on the contrary, it was conceived as an effective way to improve fuel's nebulisation and mixing achieved at the injector. We therefore addressed the study to understand whether it was possible to trigger a conformational transition for a simple aliphatic molecule, by altering its free energy landscape with the addition of a fictitious potential, ideally representing the attraction between unspecified aromatic substituents, positioned at the molecule's extremities. In this context, we observed the feasibility of the mechanism for a range of attractive energies much higher than that expected, for example, for monocyclic aromatic compounds.

In order to select a suitable aromatic moiety to test this mechanism with a more accurate model and for the subsequent experimental validation, we were required to quantify with precision the binding energy of PAH dimers, for various molecular sizes. Besides the interest raised within the spring molecule project, the focus for aromatic substituents was further motivated by an important technological issue, which outlined the framework of the second project.

Large and insoluble aromatic impurities in the form of nanometric particles, in fact, are frequently found in hydrocarbons during operations (extraction, refining etc.). Their presence is highly undesirable since are known to lead to the formation of carbonaceous deposit onto metal surfaces: importantly, they have a detrimental effect on fluid rheology and machine operations, causing a significant increase of viscosity [85] and the wear of mechanical components [86]. In addition, polyaromatic particles cannot easily be separated from hydrocarbon mixtures as they tend to clamp together and precipitate. For this reason, the synthesis of compounds aimed at preventing aggregation and the subsequent adhesion onto metal surfaces is an intense field of research. In this direction, of fundamental importance is the functionalisation of molecules with chemical groups able to bind steadily to the polyaromatic particles' surface, thus enabling the formation of a capping layer in direct contact with the solution, which can modify their

solubility and prevent noxious aggregation. Since real polyaromatic particles are composed in large majority by carbon atoms with a ratio of $sp2$ carbons higher than 50% [87], aromatic interactions are the primary candidates; an adequate chemical composition of the molecules' backbone would allow, at the same time, to tune the adsorbates's solubility according to the precise composition of the hydrocarbon mixture in which they are formed.

3.1.1 Polycyclic Aromatic Hydrocarbons

Polycyclic aromatic hydrocarbons (PAHs) are a class of organic molecules consisting of hydrogen and carbon atoms arranged in conjugated aromatic rings. These molecules are found in fossil fuels, and they can also be created through the incomplete combustion or cracking of hydrocarbons. They are non-polar compounds insoluble in water, and due to delocalisation of electrons around the aromatic rings, they are planar molecules. Because of the stability of the aromatic bond, PAHs are highly persistent molecules which are resistant to spontaneous degradation mechanisms [88, 89]; their concentration is, in fact, often used as a measure of the level of contamination of soil and air. In the last two decades, PAHs have attracted a great deal of negative attention as pollutants [90, 91, 92], and for the well-documented carcinogenic effects of the larger PAH molecules [93, 94, 95]. On the other hand, graphene has received a huge amount of largely positive press [96, 97] from academic researchers and from mainstream media because of its peculiar electronic structure.

Owing to the aromaticity of graphene and PAH molecules, it is to be expected that these interact weakly via vdW dispersion forces. As mentioned in sec. 1.4, this type of interactions is not taken into account by the standard formulation of DFT: the exchange and correlation potentials, in fact, do not treat properly the contribution arising from long-range dispersion forces, for which full ab initio post Hatree-Fock methods are required. In general, expensive quantum Monte Carlo approaches (CCSD) or perturbative methods (MP2 or higher) are expected to yield the most accurate estimation for these forces. In particular, CCSD is *de-facto* considered a benchmark, particularly in all those cases where experimental values fail to provide useful terms of comparison. However, the high accuracy of these methods requires expensive calculations that severely limit the system size to a few tens of atoms at most.

Paramount effort has been dedicated in recent years to implement various correction schemes to DFT functionals, enabling an approximated treatment of dispersion forces: DF-DFT [46] showed, indeed, to improve significantly the consistency of the binding energy for a broad series of weakly interacting molecules and ions with respect to the aforementioned MP2 and CCSD level calculations [98]. More recently, the optPBE-vdW [47] exchange-correlation functional

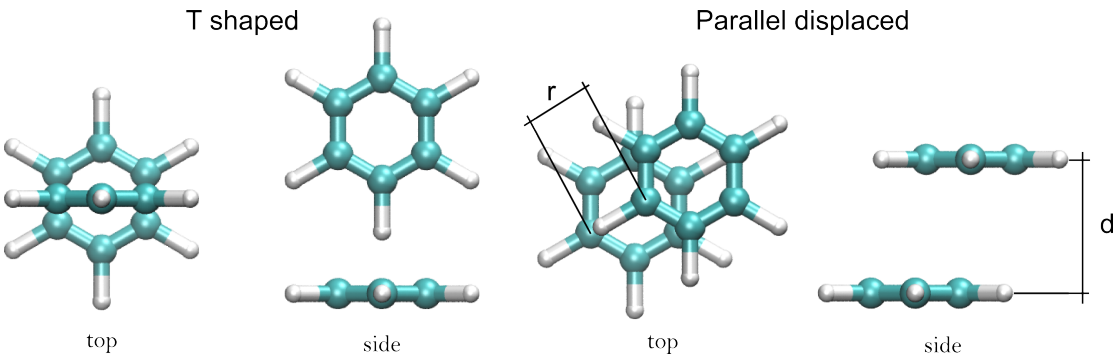


Figure 3.1: The first two minima for the benzene dimer: the *T-shaped* configuration (left) and the *parallel displaced* one (right). From [100] we have r = 1.398 Å and d = 3.641 Å.

provided a further improvement of the performances of DFT.

Among all the dispersion forces, particularly significant are those formed between parallel aromatic rings. The importance of the so called $\pi-\pi$ *stacking* interactions is due to their ubiquitous role in influencing the stability of many organic molecules (e.g. DNA) and in the formation of metal complexes [99]. For this reason, $\pi - \pi$ interactions have often received a special attention in the literature and several works have reported on the calculation of the binding energy of small aromatic compounds with metal surfaces, ions and other species. The agreement of the results, though, does not provide a clear and consistent picture, particularly for the case of benzene dimers.

Two are the main binding configurations evidenced for benzene dimers. In the first one (left of fig. 3.1), called *T-shaped*, two benzenes lie on orthogonal planes, with an hydrogen of one of the two rings placed above the centre of the other ring: the attractive interaction is exerted between the hydrogen and the sp2 carbon atoms, resulting in the aforementioned $ch - \pi$ bond. A second configuration (right of fig. 3.1), called the *parallel displaced,* results from the parallel stacking of two benzenes at a distance of about 3.64 Å and an horizontal (labelled with r in fig. 3.1) displacement of 1.40 Å. CCSD(T) level theory calculations available in literature show that *T-shaped* configuration has a binding energy 10 meV lower with respect to the parallel displaced dimer, for which values between -0.110 and -0.130 eV were reported independently in [69, 101, 102]. MP2 calculation, on the other hand, evidenced to yield higher binding energy, up to -0.170 eV [100]). The results of DFT showed, instead, to strongly depend on the precise vdW functional used: in general semi-local PBE functional with the Grimme vdW correction (DFT-D2) [103] provided results within the same range of energies evidenced by coupled cluster calculations. Interestingly, though, the T-shaped configuration was not evidenced to be the most favourable one by any of the DFT functional.

In the attempt of determining the best possible functionalisation for dispersing molecules, we modelled the interaction of PAHs of different sizes through two sets of systems: PAH dimers and PAH at graphene. This characterisation is intended to compare the binding energies of aromatic compounds of increasing dimension against two different substrates, employing the recently developed optPBE-vdW functional. Only the parallel displaced binding mode for the whole set of PAHs is considered, since the stability of the T-shaped binding geometry is expected to decrease quickly as with the size of larger PAHs.

3.2 Binding energy for PAHs

3.2.1 Computational details

The interaction energy for the series of PAH dimers and for the PAHs at the graphene surface have been calculated through DFT using the VASP ab-initio simulation package. The optPBE-vdW exchange-correlation functional was chosen since, as mentioned in previous section, it demonstrated to improve by a great amount the accuracy in the treatment of dispersion interactions for a benchmark set of weakly interacting non-polar dimers [47]. The PAH molecules considered in this study are shown in fig. 3.2: the molecules have increasing size, from benzene, consisting of just a single aromatic rings, to coronene, formed by the conjugation of seven rings.

The calculation of the binding energies for the dimers was performed as follows. Firstly, each individual molecule was relaxed alone in the gas phase. The starting configuration for each dimer was subsequently built by stacking the minimised structure of two identical PAHs in parallel configuration at an initial distance d of 3.7 Å, with a displacement r of about 1.4 Å as labelled in fig. 3.1. This initial guess was chosen using the optimised geometry of benzene dimers according to [104]. The binding energy for PAH dimers was then calculated as the difference between the minimum energy of the dimer system and the energy of each molecules relaxed alone counted twice. The binding energy of PAH at the graphene surface were similarly calculated, after having relaxed the graphene sheet separately.

The size of the unit cell was set for each system to an adequate value, leaving about 10 Å of vacuum padding in each direction; this spacing was shown to be sufficient to prevent the interaction with periodic replicas of the system. A cut off of 450 eV was used for the kinetic energy. The geometry optimisation was performed using a conjugate gradient algorithm, with a convergence criterion for ionic relaxation of the forces of 0.05 eV/Å. The sampling of the Brillouin zone was limited to the $\mathbf{k} = \mathbf{\Gamma}$ point for dimers; for the case of graphene, the periodicity of the honeycomb lattice required to expand the sampling to a 6x4x1 grid of k points. Since

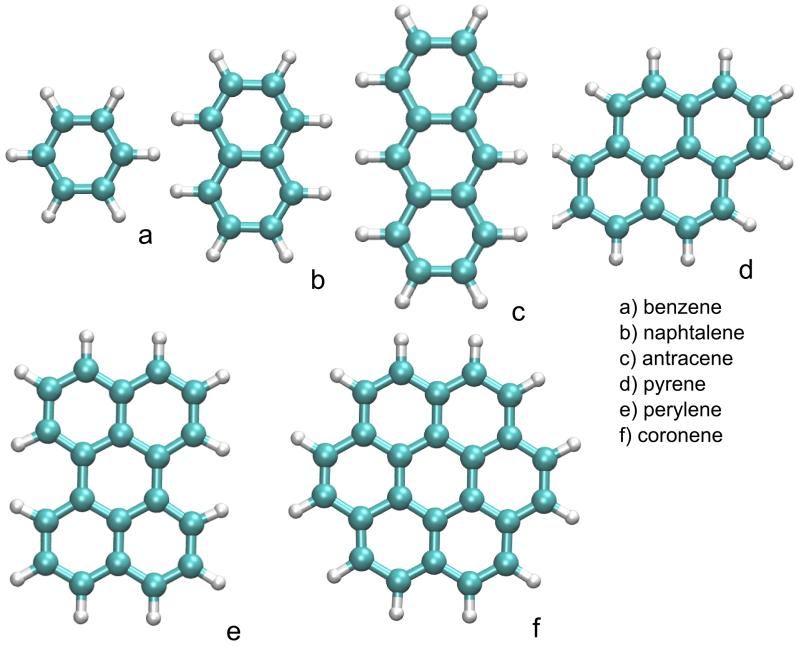


Figure 3.2: Polycyclic aromatic hydrocarbons (PAHs) considered in this study.

the molecules of fig. 3.2 have considerably different sizes, a 6x6 carbon atom supercell was chosen for benzene (12.8 Å of width), while a 8x8 cell (17.1 Å of width) was necessary for the larger PAHs; a 12 Å of vacuum was left in the direction orthogonal to the surface to avoid the interaction with replicas.

3.2.2 Calculated energies

The binding energies for PAH at graphene are reported in the second column of table 3.1. As expected, the energy was found to decrease for an increasing molecular size, analogously to what was already reported in other studies [105, 104, 69]. Values of -0.57 eV and -0.88 eV have been calculated for benzene and naphtalene, against experimental measurements of 0.50±0.1 eV and 0.85±0.1 eV, respectively [106]. The agreement is less good for the case of coronene, for which a value of -1.779 eV was calculated, against an experimental estimation of 1.5±0.2 eV. The comparison with the available computational references was somewhat biased by the use of different functionals. Chakarova *et al.* [107], for example, using the revPBE-vdw functional reported physisorption energies of -0.50 eV and -0.76 eV, which are about 10% higher than the results obtained by our calculations.

The binding energies for the dimer systems are reported in the second column of table 3.2 for the whole series of PAHs; only for three of them it was possible to find comparable reference

values in the literature. We estimated binding energies of -0.15 eV, -0.34 eV and -1.03 eV for the benzene, naphtalene and coronene dimers, respectively. Also in this case, the results obtained are 10% lower than those reported by available recent DFT calculations: Chakarova *et al.* [69], found binding energies of -0.12 eV for benzene and -0.29 eV for naphtalene with a revPBE-vdw functional; -0.98 eV were reported by Feng *et al.* [108] for coronene dimers with a PBE-vdw functional.

PAH-graphene	E_{DFT} $[eV]$	E_{GAFF} $[eV]$	E_{OPLSA} $[eV]$
Benzene	-0.569	-0.469	-0.351
Naphthalene	-0.878	-0.761	-0.581
Antracene	-1.145	-1.054	-0.815
Pyrene	-1.282	-1.183	-1.083
Perylene	-1.4883	-1.183	-1.335
Coronene	-1.779	-1.750	-1.538

Table 3.1: Binding energies for PAHs at graphene calculated with DFT (second column) and the respective energies obtained by the minimisation of two classical potentials (this and fourth columns). Errors are within the DFT accuracy level of about 0.01-0.02 eV.

PAH dimer	E_{DFT} $[eV]$	E_{GAFF} $[eV]$	E_{OPLSA} $[eV]$
Benzene	-0.152	-0.100	-0.160
Naphtalene	-0.344	-0.151	-0.339
Antracene	-0.526	/	/
Pyrene	-0.622	/	/
Perylene	-0.832	/	/
Coronene	-1.031	-0.621	-1.278

Table 3.2: Binding energies for PAH dimers calculated with DFT and the corresponding energies calculated through the minimisation of two classical potentials. Errors are within the DFT accuracy level of about 0.01-0.02 eV.

These findings provide a set of energies that characterise the $\pi - \pi$ stackings of PAHs, producing a useful table relating the size of the aromatic compound with its corresponding binding energy for the two binding types. Additionally, the DFT calculations suggested that optPBE-vdW functional slightly overestimates the contribution of aromatic interactions when compared with analogous data obtained with other functionals, although the comparison of the binding energies for a large set of dimers demonstrated how the performances of optPBE-vdW are higher than those of most vdW functional, including the revPBE-vdW [109, 47].

In order to compare the binding of different PAHs for the two sets of systems, the energy of tables 3.1 and 3.2 needed to be normalised. The number of carbon atoms was chosen as a normalisation criterion, being a more natural and straightforward choice respect to, for example,

the number of aromatic rings. The normalised energies are plotted in fig 3.3, as a function of the number of carbon atoms of each PAH molecule. Interestingly, the two curves present a non-linear dependence and approach the asymptotic limit represented by the graphite's exfoliation energy, which, according to our calculations, is equal to -0.052 eV per carbon atom (in good agreement with [110]). Importantly, these plots allow one to observe two opposite trends. On one hand, the normalised energy for PAH dimers decreases for an increasing number of carbon atoms: a value of -0.02 eV was calculated for benzene, while an energy -0.04 eV per carbon atom was found for coronene. Conversely, the normalised physisorption energy at graphene was observed to get progressively weaker for PAH of increasing size: a binding energy of -0.095 eV per carbon atom was determined for benzene, which was found to increase by nearly 30% for the case of coronene, for which we report an energy difference of -0.074 eV.

As already brought forward by other works [108, 104], it was expected to get relatively more favourable binding energy for larger dimers, because of the higher fraction of sp2 carbon atoms with respect to the number of hydrogen. However, to our knowledge, the combination of these to opposite trends have never been highlighted before. Importantly, the smaller the PAH molecule, the more favourable is the binding energy with a graphene surface, here representing a general aromatic substrate. This finding has an interesting implication: the total binding energy resulting from the interaction of several smaller PAHs with a big aromatic substrate is lower than the one resulting from using larger PAHs. For example, four benzene molecules are expected to bind more strongly (cfr. energy in table 3.1 resulting in E=-0.497 eV) at graphene than a single coronene molecule, despite having the same total number of carbon atoms. This is an important result that was never observed before and paves the way for the rest of this study.

3.2.3 DFT-based correction to the classical potential for $\pi - \pi$ interactions

In the last section, we presented DFT calculations of the binding energy between PAH dimers and PAHs at graphene. DFT, although generally accurate, is too slow to run extensive MD simulations. Classical potentials can facilitate larger MD simulations, but the accuracy of the potential used needs to be checked first. To do so, the same energy calculation were performed with two widely used *ff*.

The third and fourth columns of table 3.1 display the binding energies of PAH molecules at graphene calculated with the GAFF [111] (E_{GAFF}) and the OPLS-AA [112] (E_{OPLSA}) *ff*, respectively. The agreement of the binding energies determined via the selected classical potentials was not satisfactory overall. The GAFF *ff* reproduced the energies of PAHs at graphene with

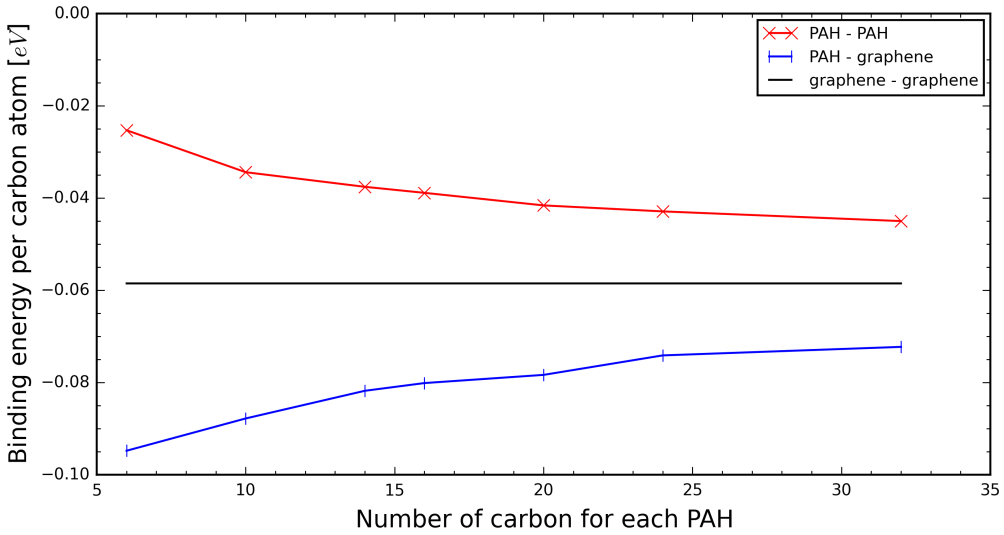


Figure 3.3: The binding energy per carbon atoms for the PAHs listed in fig. 3.2 is plotted for the dimer systems and the PAH at graphene. The physisorption energy per carbon atom for PAHs at graphene (blue line) gets weaker with the increasing size of the molecules, whereas the binding energy of PAH dimers (red line) gets stronger for increasing PAH size. Both binding curves approach asymptotically the graphene-graphene exfoliation energy, which is shown here as a solid black line.

an average difference of about 10%, while the binding energy for PAH dimers were largely lower estimated: particularly, an energy of -0.151 eV was reported for naphtalene against a DFT value of -0.344 eV. Conversely, OPLS-AA reproduced the energy of PAH dimers more closely: values of -0.160 eV and -0.339 eV were calculated for benzene and naphtalene, which compared well with the DFT results. However, the agreement of OPLS-AA energies for PAHs at graphene is less satisfactory, since the physisorption energies are off by above 30%. Clearly, neither the two *ff*s allows an accurate and transferable representation of aromatic interactions, since in both cases at least one of the series of energies was approximated by an unacceptable amount.

To obtain reliable results via MD calculations, one of two classical potentials needed to be corrected to realign the energy values. The first possibility examined was to producing a new fully parameterised in-house force field for aromatic rings capable of fitting the DFT data: this would be expected to substantially improve the accuracy of the model at the cost of a systematic increase in complexity and development time. The alternative was to introducing a corrective term into one of the two force-fields. Since the binding energies were shown to follow the correct trend, the second strategy was preferred. We therefore added a new attractive term applied between the aromatic carbons, in order to correct the GAFF potential and yield quantitative results comparable with the DFT ones. This *ff* was chosen since it provided a better agreement for the binding energies with a large substrate, thus best suiting the needs of this study and

guaranteeing, at the same time, consistency with the parametrisation of hydrocarbon molecules that will be presented later in this chspter. The additional term was required to leave all the other interactions unmodified, in particular, those between atom pairs not belonging to aromatic groups. Different schemes for this correction have been considered, which progressively improved the agreement with DFT results.

Correction schemes

For a pair additive potential, the modification of the parameters for non-bonding interactions of a particular atom type would affect the interactions with all the other atoms, since the mixing rule is routinely adopted (see footnote on pag. 16). To prevent this, the mixing rule does not have to be applied but the all parameters have to be assigned singularly. A simple strategy for the correction of $\pi - \pi$ interactions can be introduced by just modifying the non-bonding terms of the potential between aromatic carbons. For the first correction scheme attempted, a simple 12-6 LJ term was added resulting in the following expression

$$V_{add}^{(a)} = \left\{ \frac{q_{ca}^2}{4\pi\epsilon_0 r_{ij}} + 4\epsilon_{ca} \left[\left(\frac{\sigma_{ca}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ca}}{r_{ij}} \right)^{6} \right] \right\} + 4\epsilon_a \left[\left(\frac{\sigma_a}{r_{ij}} \right)^{12} - \left(\frac{\sigma_a}{r_{ij}} \right)^{6} \right], \tag{3.1}$$

where r_{ij} is the distance between the aromatic carbon i and j, while σ_{ca} and ϵ_{ca} are the original parameters of the LJ part of the GAFF potential pertaining to aromatic carbons; the last term on the right hand side of eq. equation (3.1) was added only between two aromatic carbons, where σ_a and ϵ_a are the parameters of the first corrective scheme. The determination of the parameters for the added term was carried out by fitting the DFT potential energy profiles for benzene dimers with the corresponding energy profile obtained by the optimal choice of σ_a and ϵ_a of the corrected classical potential. The best value for σ_a was found to be equal to 2.94 Å, while ϵ_a corresponded to 0.039 eV. Using these parameters, also the binding energy of naphtalene displayed a good match (-0.343 eV compared to a DFT binding energy of -0.344 eV); however, the binding energy for the coronene dimer resulted greatly lower-estimated, since an energy equal to -1.681 eV was obtained against a value of -1.031 eV calculated with DFT. This discrepancy might originate from the fact that a decrease of the electronic density around the geometric centre of the coronene molecule was observed with DFT, while this cannot be taken into account accurately with a classical potential without defining additional atoms types depending on their position within the molecule. Another observation worth making concerns the large difference between the distance parameter σ_a, for which a value equal 2.94 Å was determined, and the benzene dimer separation distance of 3.65 Å: this difference showed that the effect of this proposed correction scheme is to aligning the energy profiles by pulling against

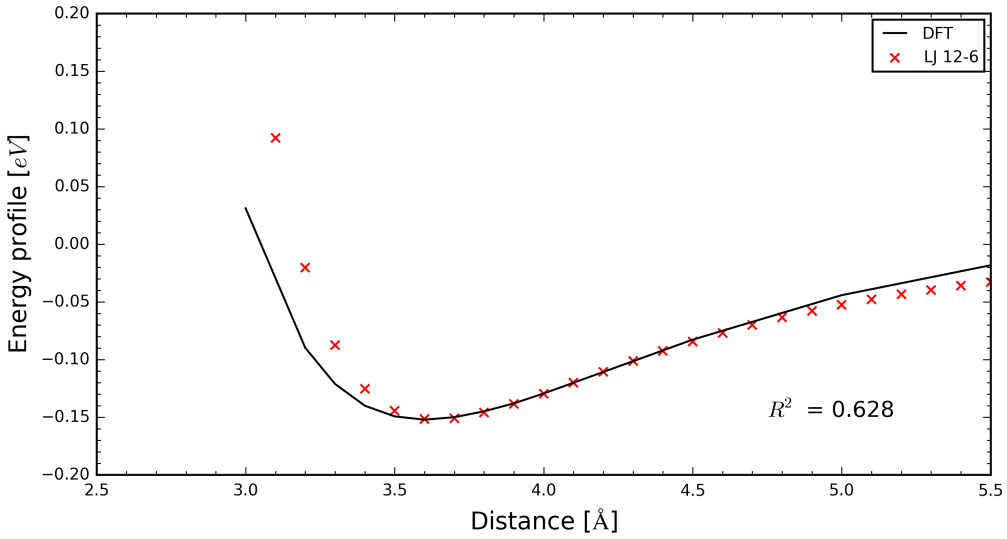


Figure 3.4: Comparison of the benzene dimer binding energy profile in the parallel-displaced configuration computed with the $V_{ca}^{(b)}$ scheme (red points) with the DFT binding energy curve (black line); the R^2 value is reported to give an indication of the sum of residues.

the repulsion exerted by the original vdW and electrostatic terms (first and second terms of eq. 3.1).

The latter consideration suggested to adopt a second parametrisation for the corrective term. The potential energy term was set to the following,

$$V_{add}^{(b)} = 4\epsilon_b \left[\left(\frac{\sigma_b}{r_{ij}} \right)^{12} - \left(\frac{\sigma_b}{r_{ij}} \right)^{6} \right]. \tag{3.2}$$

In this case, no electrostatic term was included, consistently with the idea that electrostatic repulsion between partial charges assigned to carbon atoms might be spuriously lowering the dimer binding energies. The value of the ϵ_b parameter for the benzene-dimer was found to be 0.006 eV. Using these values, the binding energies of the naphthalene and coronene dimers corresponded to -0.339 eV and -1.278 eV, respectively. Importantly, the agreement of this second correction scheme with DFT energies (-0.343 and -1.031 eV) was significantly improved. However, the binding energy of the coronene dimer was still overestimated by non-negligible amount, although improving considerably the agreement with respect to the first scheme (see table table 3.3).

The binding energy profile for a benzene dimer is reported in 3.4 as a function of the inter-plane distance d to assess the match between the resulting classical energy profile and the DFT one. Clearly, the 12-6 LJ potential allowed to recover the minima energy in correspondence of the correct separation distance as well as the attractive long range part of the profile. However,

the repulsive portion of the curve evidenced to be too stiff, having a significantly different slope leading to the formation of a narrower basin; this resulted in an unsatisfactory fit, expressed by the low R^2 value of 0.628 (here intended to be representative of the sum of the residues). To overcome this issue a weaker repulsive term was, therefore, needed.

PAH-PAH	E_{DFT} $[eV]$	$E^{(a)}_{\sigma,\epsilon}$ $[eV]$	$E^{(b)}_{\sigma,\epsilon}$ $[eV]$	$E^{(c)}_{\sigma,\epsilon}$ $[eV]$
Benzene-benzene	-0.152	-0.152	-0.156	-0.154
Naphthalene-naphthalene	-0.344	-0.343	-0.356	-0.344
Coronene-coronene	-1.031	-1.681	-1.278	-1.145

Table 3.3: PAH-dimer binding energy values obtained via DFT calculations and the corresponding energies for the three correction schemes.

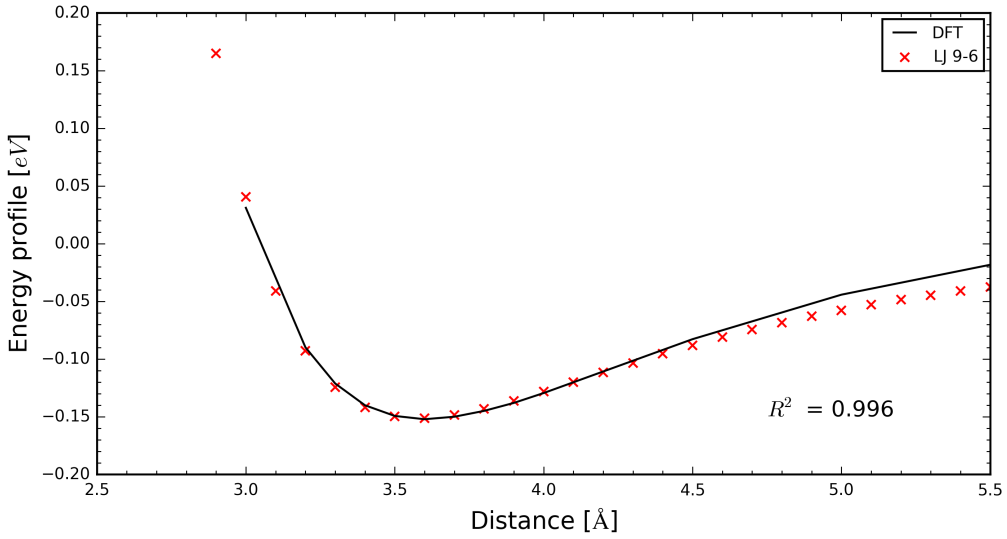


Figure 3.5: Comparison of the benzene dimer binding energy profile in the parallel-displaced configuration computed with the $V^{(c)}_{ca}$ scheme (red dots) with the DFT binding energy curve (black line); the R^2 value is reported to give an indication of the sum of residues.

$$V^{(c)}_{ca} = 4\epsilon_c \left[\left(\frac{\sigma_c}{r_{ij}} \right)^9 + \left(\frac{\sigma_c}{r_{ij}} \right)^6 \right] \tag{3.3}$$

The 9-6 Lennard Jones potential shown in eq. equation (3.3) was than chosen to improve the fit. The parameters $\epsilon_c = 0.0087$ eV and $\sigma_c = 3.4$ Å were determined by fitting the benzene dimer energy profile to the reference one as before. This new scheme gave a very good agreement for the PAH dimer binding energies of benzene and naphthalene, reducing the difference with the energy of coronene dimers to just 10% (see third column on table table 3.3). Importantly, the match of the binding energy curves (see fig. 3.5) was greatly improved, at the cost of a slight

bias in the attractive portion for a distance greater than 4.5 Å. The value of R^2 value was much improved respect to the second parametrisation, being equal in this case to 0.996. To further test this correction scheme for other PAHs, the physisorption energies at graphene of benzene and naphtalene were calculated: in the first case, an energy of -0.570 eV perfectly matched the -0.569 eV calculated with DFT; for the second case, an energy of -0.905 eV was found, again, in good agreement with the -0.878 eV calculated with DFT.

Additional prove of the reliably of this third parametrisation, was found in an earlier computational study [77], in which a 9-6 LJ potential was determined to be the best possible choice among several tested analytical expression for a classical potential fitting of ab-initio energy profiles of a similar set of aromatic dimers. For the investigation presented in the remainder of this chapter, this last correction scheme was used for the interaction between aromatic groups.

3.3 Benzene-terminated dendron

The results in fig. 3.3 suggest that, in order to maximise the interaction with a graphene-like substrate, smaller aromatic compounds are expected to bind more strongly than fewer larger ones. This concept is particularly important for the engineering of molecules able to bind steadily to large aromatic compounds. A similar design strategy is reminiscent of one of the main characteristic of dendritic molecules, a particular class of compounds known for their multivalent recognition properties, which have been used, in particular, for the interaction with biomolecules. Dendrons and dendrimers, thanks to their highly branched structure, can accommodate a high density of functionalisations, thus providing a flexible platform for many host-guest applications. For this reason, dendron molecules have attracted in recent years a great deal of attention across multiple disciplines for their potential applications, spanning from surface chemistry [113, 114, 115, 116] to drug delivery [117, 118, 119]. In this case, the higher binding efficacy expected for smaller PAHs is an ideal example in which dendritic structures are a particularly suited molecular model.

As proof of concept for the feasibility of the above design strategy, a simple dendritic structure (see fig. 3.6) was conceived as a test case. A model dendron molecule consisting of three aliphatic legs (12 and 6 carbon atoms long), each terminated by a phenyl substituent was created and parametrised using the GAFF *ff*, together with the correction described earlier applied to the aromatic carbons (the partial charges were calculated with the semi-empirical method AM1-BCC [120]). The aim was to verify the efficacy of small aromatic terminations as anchoring points for a stable physisorption onto a graphene sheet, in presence of competing interactions,

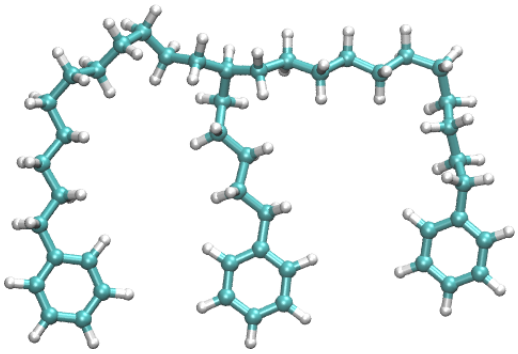


Figure 3.6: Molecular structure of the model dendron molecule: each of the aliphatic leg is terminated with a phenyl group in order to exploit the multivalent interaction with graphitic substrates.

such as those ones exerted by (i) aromatic terminations with themselves, (ii) solvent molecules and (iii) topological defects of the surface.

The energies reported in fig. 3.3 suggest that a cluster involving several dimerised benzene molecules, upon coming into contact with a graphene sheet, should spread out to bind at the sheet. This mechanism was tested with the dendron model. Firstly, the molecule was minimised in the gas phase. As anticipated, the phenyl terminations came closer due to the attractive vdW interactions - see the closed-structure in fig. 3.7 a. The relaxed structure was further minimised by DFT with the optPBE-vdW functional: the relaxation converged within few iterations, providing further proof of the accuracy of the classical potential combined with the corrective term compared to higher precision methods. The DFT total binding energy difference of the closed dendron structure calculated in gas phase was 0.97 eV lower than open dendron structure, part of which is expected to account for the interaction between the three phenyl groups.

The relaxed dendron structure was subsequently placed above a graphene sheet modelled by a 20x20 unit cells, with the phenyl group positioned at 5 Å from the surface. Since the graphene sheet was meant to act just as a rigid substrate, its carbon atoms were kept fixed with harmonic constraints; the dendron-graphene system was then simulated via MD at 300 K for a total of 300 ps. The dendron structure quickly opened with two of the three phenyl rings which readily bound to the graphene sheet within the first $\backsim$30 ps of the simulation. The molecule then stayed anchored to the substrate for additional $\backsim$40 ps before the third termination also lied down to the surface in a parallel configuration. The dendron remained attached to the surface for the remainder of the simulation (about 250 ps). Several different relaxation were also perfomed using different starting configuration exhibiting a ready binding of the phenyl moieties in an analogous way.

Three snapshots taken from one of the trajectories are shown in fig. 3.7. The phenyl groups were

observed to bind at the surface in a nearly flat configuration, where the tilt is an indication that the $\pi - \pi$ stacking has been slightly perturbed by the steric hindrance caused by the aliphatic portions of the molecules joining these phenyl groups. This was observed to give rise to a negligible (typically less than 3%) reduction in the phenyl/graphene binding energy estimated at DFT level.

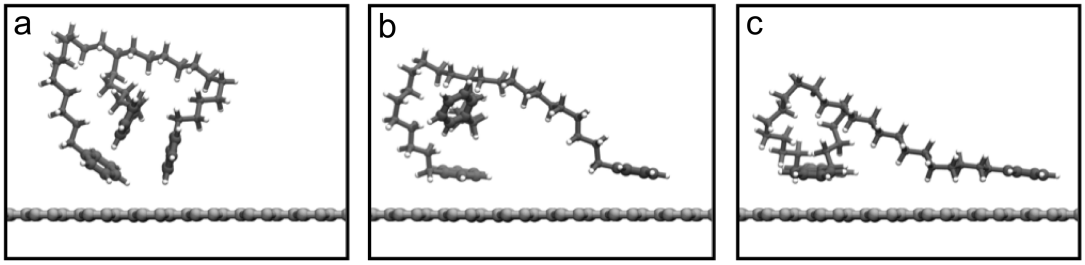


Figure 3.7: Three snapshots taken from the simulation of a model dendron at graphene. The dendron starts moving from the minimised closed configuration placed over the graphitic surface (a): the phenyl-terminated legs are bound to each other. Two of the three phenyl end-groups interact with the graphitic surface (b), and eventually all three phenyl rings bind to the surface (c) .

With this simple set of simulations using classical potential and more accurate DFT based molecular dynamics, we demonstrated that smaller PAH functionalities binds readily to a model aromatic substrate, and that the preference for PAH-graphene binding over PAH-dimers leads to gas phase PAH clusters opening up in the presence of a suitable binding site (here represented by a graphene sheet). This is a very interesting results from a chemical point of view and, more importantly, for the possible technological applications. In the specific case, this dendron model represents a promising concept for the design of molecules capable of interacting selectively with large heterogeneous aromatic compounds. As mentioned earlier, this responds to an important technological issue, particularly for the case of those undesirable particles which are frequently found suspended in hydrocarbon mixtures. The formation of aliphatic layers onto polyaromatic particles would be, in fact, an effective method to increase their solubility.

3.3.1 Model the substrate/solvent interface

To extend the results of previous section, a more realistic model system was required to verify the ability of the proposed dendritic structure to stick onto polyaromatic substrates. Clearly, the simulation of the dendron in the gas phase at the clean graphene surface, while supporting the feasibility of the mechanism, did not aim at representing the complexity of the physisorption onto a real polyaromatic particle dispersed in hydrocarbons. Firstly, a clean graphene surface does not render adequately the irregularity of spontaneously formed particles, which are composed

by heterogeneous chemical groups and defective aromatic areas. Studies conducted on the microstructure and composition of carbonaceous deposits, indeed, revealed how are constituted by the agglomeration of inhomogeneous, nanometric-sized particles of 2-5 nm of diameter, each of which is formed by interlaced bundles of graphene sheets [87]: the continuous bending of the graphene segments evidenced by high resolution TEM images, confirmed the presence of sizeable fractions of sp3 carbons (about $46\%_w$). A high concentration of heterogeneous functional groups was also measured by XPS spectroscopy, which allowed to determine an oxygen content of about $16\%_w$: among the major components of the spectra the contribution of several carbon-oxygen functional groups, such as C=O, C-OH and C-O-C, was detected. The high concentration of polar groups substantiates the hydrophilic behaviour of polyaromatic particles.

Given the complexity of these substrates, we did not focus on their full characterisation from a chemical point of view, for which little is known on the atomistic scale. Rather, the attention was drawn onto mimicking a certain level of topological irregularity, in order to investigate how steric hindrance effects associated with a highly heterogeneous surface might affect the balance between dendron and solvent sticking. To this end, we conceived a model surface based on a graphene sheet modified with the insertion of methyl groups on randomly chosen sites, at two different surface coverages, for comparison. Methylation of graphene, while providing a simple representation for a defective aromatic surface, has been shown to occur at operative condition by the reaction with highly reactive radicals [121], which, in turn, are abundant in hydrocarbon mixtures exposed at high pressures. The objectives were (i) to establish whether the dendron can bind to a polyaromatic substrate in the presence of the solvent, and if so, (ii) to identify whether the dendron is capable of "adapting" to the surface's irregularities by virtue of its molecular structure, and anchor itself securely to the surface equally well, or possibly better than the solvent molecules. A clean graphene (*FLAT*), a moderately methylated (*SPARSE*) and highly methylated surfaces (*DENSE*) are shown in fig. 3.8 (subfigures a, b and c respectively). The methyl groups, displayed in red, significantly reduce the dimension of the aromatic regions, while having very little impact on the surface chemistry.

3.3.2 The hydrocarbon model

The second important element for the creation of a realistic model system was the selection of a solvent that could adequately reproduce the chemical and physical properties of the reference fluid. An example of the composition of a standard low-sulphur diesel is reported in fig. 3.9: the table and the bar plot show how diesel is composed of a large variety of aliphatic and aromatic compounds of different sizes, with branched and linear alkanes constituting the larger fractions.

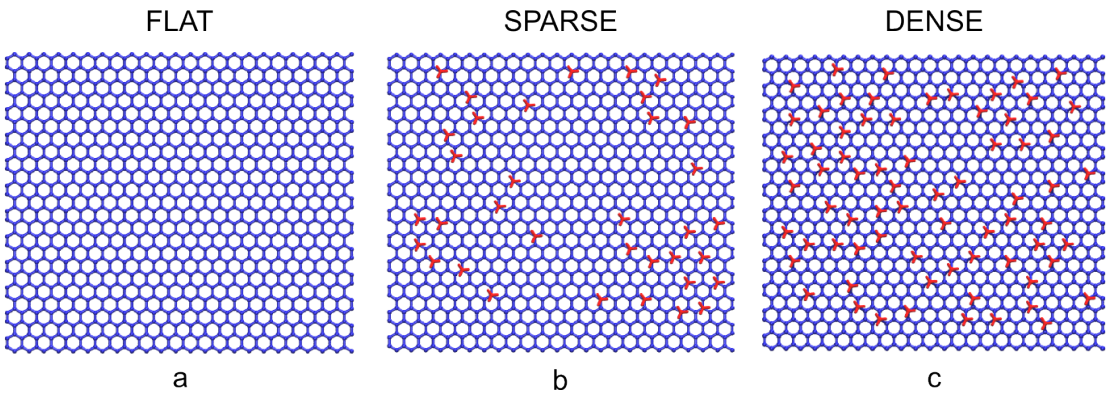


Figure 3.8: The three different model surfaces: (a) the *FLAT* graphene surface; (b) a *SPARSE* methylated graphene surface (35 methyl groups); (c) a *DENSE* methylated graphene surface (70 methyl groups). The graphene carbon atoms are rendered in blue while methyl groups are coloured in red.

Compounds	Diesel composition $[mg\ g^{-1}]$
n-Alkanes	190.7
Branched alkanes	368.4
Saturated cycloalkanes	52.8
PAHs	1.8
Alkylated PAHs	25.3
others	44.8

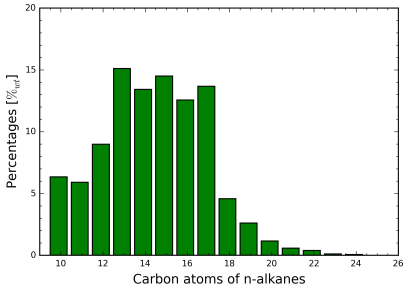


Figure 3.9: A low-sulphur diesel reference composition is reported, listing the fractions of the major molecular types (left); the n-alkanes fraction contains several compounds into having a different number of carbon atoms (right) [122].

The most important species within the second group is *n-hexadecane*: n-hexadecane is, in fact, one of the most abundant molecules and its concentration is often used as an index for the combustion properties of commercial diesel fuels, since it is known to yield optimal combustion performances. The choice of n-hexadecane as model solvent, was also supported by few additional considerations. Firstly, hexadecane molecules matched in the size and composition the backbone of the dendron, thus guaranteeing an adequate solubility and a comparable mobility. These were two important factors, since they are expected to represent the condition for which the binding of the dendron to the surface is not thermodynamically or kinetically favoured over the adsorption of the solvent molecules. Moreover, the choice of a solvent model constituted by a single compound, in place of a more complex mixture, made easier the comparison with available experimental results during the parametrisation procedure.

Few preliminary MD runs have been carried out to determine the equilibrium properties of a hexadecane bulk system for two different *ffs*. The GAFF *ff* was tested first: the equilibrium

density found at 300 K corresponded to a value 0.833 g/cm^3, which was about 8% higher that the reference value of 0.770 g/cm^3 [123]. The reason for this discrepancy was found to be caused by the occurrence of the phase transition of the system from a liquid to a waxy state: the alignment of the aliphatic chains into crystalline subdomains evidenced the occurrence of the nucleation of a solid phase that was expected to take place at a lower temperature, at about 291 K. As observed in other works [124], the incorrect melting temperature and the resulting density overestimation are caused by a non-optimal parametrisation of the non-bonded and the torsional component of the classical potential (see eq. 1.3). To overcome this issue, we modified the LJ parameters to match those of the OPLS-AA *ff*, which is known to yield more accurate results for equilibrium properties of n-alkanes while still conserving transferability [125, 126, 127]; all the other parameters for the remaining terms of the potential were, instead, maintained to the original values set by the GAFF *ff*. With this modification, the calculated density of hexadecane at the same temperature was lowered to 0.781 g/cm^3, a value which is within a 1 % error margin respect to the experimental references at the same condition. The density and the bulk modulus were further calculated and compared for a wide range of pressures (1-2000 atm) and temperatures (300-473 K), confirming again good agreement with available experimental data.

The model systems were than built as follows. A simulation box having cross section in the XY plane and size corresponding to the one of model surfaces displayed earlier, was created (59.4 Å x 51.4 Å). The dendron was placed in its relaxed configuration (see fig. 3.8 a) at a distance 10 Å from each of the three surfaces. This distance matched to the PAH-PAH force cut-off, such that no dendron-substrate interaction was initially exerted. The box was then solvated with 300 hexadecane molecules, a number sufficient to adequately fill the simulation box to an estimated height of about 50 Å: this height, reached upon density equilibration, was considered sufficient to avoid the simultaneous interaction of the dendron with both the graphene surface's replicas. MD simulations of the hexadecane solvated dendron at the three model surfaces at 300 K were carried out to investigate the occurrence of sticking events. The simulation were conducted in the NVT ensemble using a Langevin thermostat (damping parameter equal to 1 ps^{-1}). For simplicity, the motion of the carbon atoms of the model surfaces was prevented through the application of harmonic positional restraints. During the first 5 ns of the simulation, the solvent was equilibrated to the bulk density, while the dendron was kept fixed at its initial position by harmonic restraints to avoid undesired early diffusion; upon equilibration, the height of the simulation box was conveniently adjusted to its final value (50 Å). This approach is similarly to that used for the simulations in chapter 4 and was preferred to conduct the simulation in the NPT ensemble to reduce the instability caused by the decoupled scaling of the pressure tensor components together with the armonic restraints imposet to the aromatic subtrate's atoms.

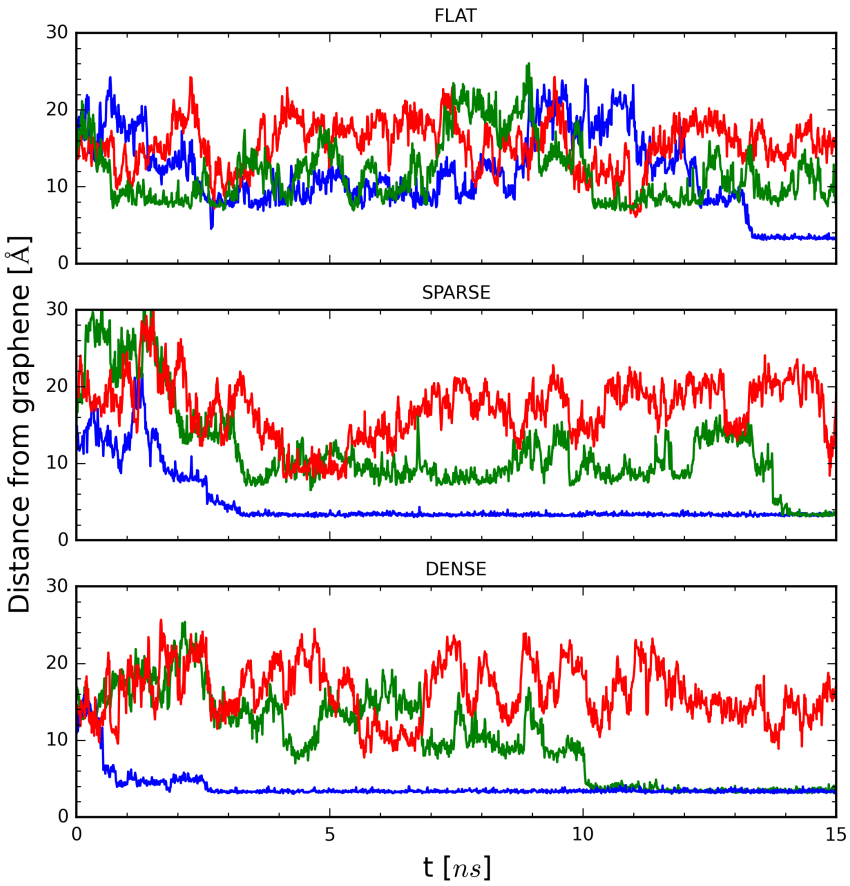


Figure 3.10: The position of the centres of mass of the three phenyl groups is plotted against the simulation times for each system. Within 15 ns, just one sticking event was observed on the *FLAT* graphene surface (top panel), while for both for the *SPARSE* system (middle panel) and the *DENSE* one (bottom panel) two phenyl groups were observed to penetrate the hexadecane adlayer and bound to the model surfaces.

The following 15 ns were then conducted removing the harmonic restraints on the dendron, while keeping other simulation's parameters unmodified; the trajectories were then used for the analysis.

3.3.3 The sticking mechanism

The analysis of the trajectories revealed that the dendron's anchoring was observed on all the surfaces considered. For all systems, in fact, at least one phenyl termination was able to stick to the aromatic substrates in a parallel configuration, finding its way through the layers of hexadecane molecules. Upon coming in proximity of the first adsorbed layer of hexadecanes, a phenyl moiety was eventually able to penetrate the first solvent adlayer and bind to an adsorption site. The observed sticking mechanism initiated from the formation of a "T-shaped" binding

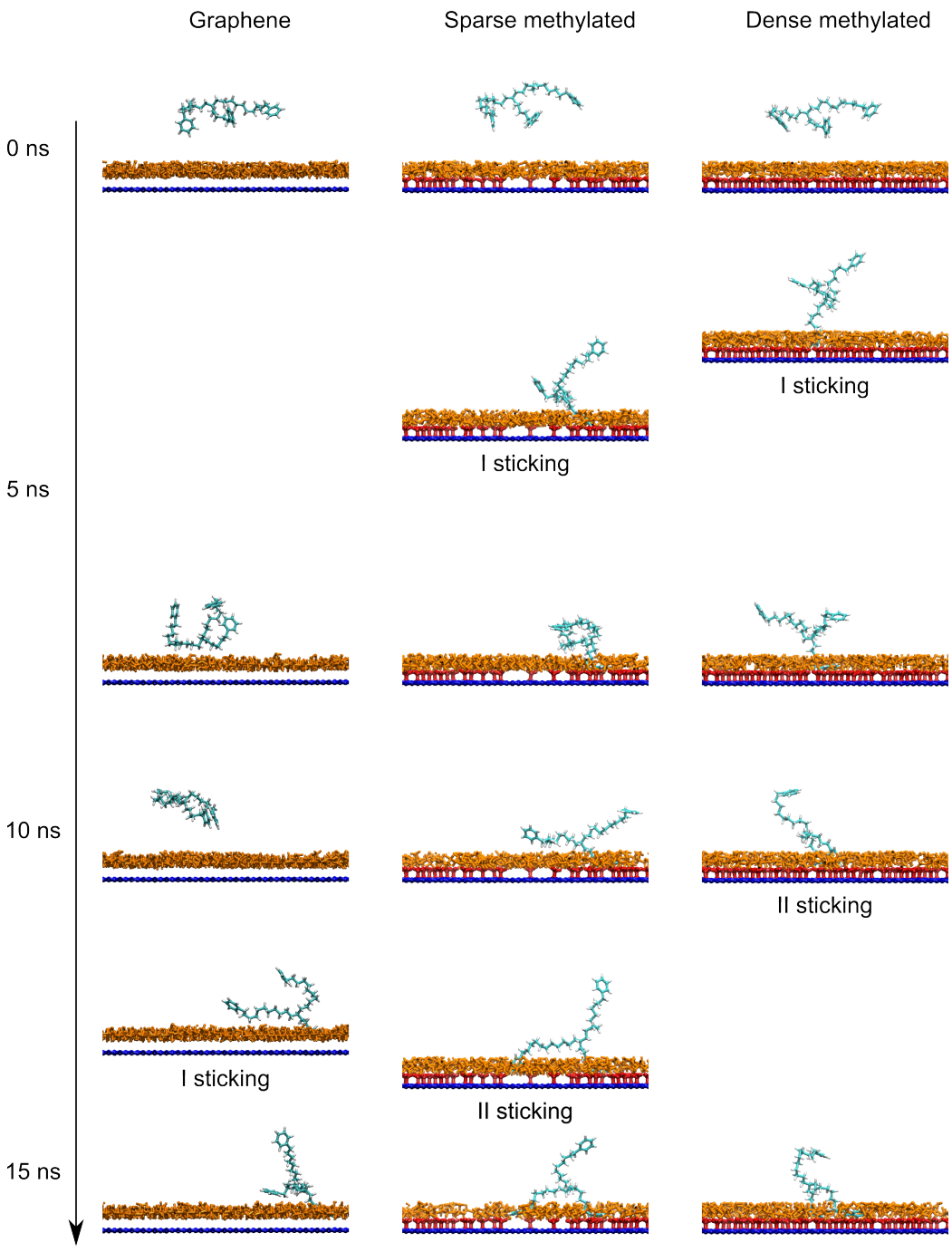


Figure 3.11: Some snapshots extracted from the trajectories show the occurrence of the sticking events onto the three model surfaces (cfr. fig. 3.10). Only the first adlayer of hexadecane molecules (coloured in orange) is rendered for clarity.

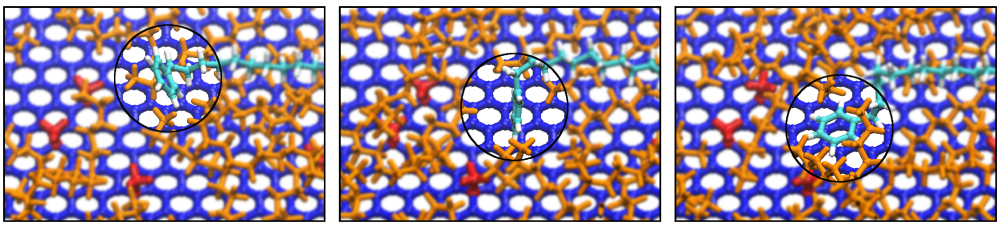


Figure 3.12: Details of the binding mechanism: a phenyl group initially occupying a position above the hexadecane adlayer (left), eventually accessed the surface in a T-shaped configuration (middle) before flattening down in a parallel alignment (right).

mode with graphene; from this position, the simultaneous rotation of the phenyl group and the replacement of the neighbouring hexadecane molecules, enabled the phenyl group to reach a stable parallel binding with the aromatic surface, which was found to be stable for the whole length of the simulation. Some snapshots taken from the MD trajectory for the *SPARSE* system are shown in fig. 3.12, illustrating three successive steps of the binding mechanism.

Despite the sticking event was detected with at least one phenyl groups at all the three surfaces, a remarkable difference was evidenced in the time needed for the first binding event to take place: the higher the fraction of defective aromatic areas, indeed, the more readily the sticking events was observed. This quite counter-intuitive finding can be monitored in fig. 3.10, where the distance between each phenyl terminations and the model surface is plotted against the simulation time for all the systems. On the *FLAT* graphene sheet (top), the three phenyl groups were found to be at a distance greater than 8 Å (thus at a distance comparable with the force cut off) for most part of the run, before one group, after about 13.2 ns, reached the surface, steadily forming a bond at distance of about 3.7 Å. The snapshots in the left column of fig. 3.11, show some renderings of the sticking process along the trajectory: the dendron spent most part of the simulation above the graphene surface before anchoring through a single phenyl group, while the remaining two terminations remained far away from the surface. For the *SPARSE* system, though, characterised by a moderate fraction of methylated sites, the first sticking event initiated after just 2.5 ns. After the first binding event, the dendron remained steadily tethered to the surface for the subsequent 11 ns, before a second group was able to approach another aromatic site, binding to it as before (14 ns). For the *DENSE* system, having twice as many methyl groups, the first event was observed after just 2 ns, while the second verified about 10 ns after the beginning of the equilibrated run.

The results of these simulations provided substantial elements to assess our hypothesis and to draw some conclusions. Firstly, the proposed dendritic molecule was proved to be a viable strategy for the interaction with complex aromatic substrates. The binding energy of a phenyl

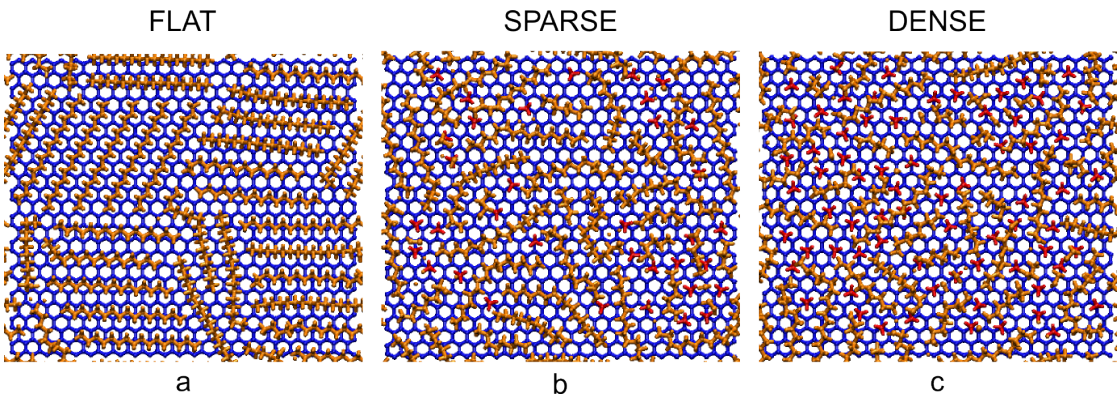


Figure 3.13: Surface ordering for the adlayer of hexadecanes on the three different model surfaces: clean graphene (a), sparsely methylated graphene (b) and densely methylated graphene (c). The packing order of the bottom layer of hexadecane molecules was shown to depend significantly on the surface topology.

group corrected at the DFT level was shown to be sufficient to form a stable binding. The multivalency effect of dendritic structure was displayed, since in two cases, two dendron terminations were able to interact with the model surface. In particular, the proposed dendritic structure proved to be effective for the interaction with defective aromatic surfaces to an extent that was not foreseen. Before sticking at the surface, the phenyl end groups must first be able to penetrate through the first layers of the adsorbed hexadecanes and find a suitable aromatic adsorption site (see fig. 3.12). This mechanism was found to be more difficult at the flat graphene surface. The different morphology of the heterogeneous models surfaces led to hypothesise that this was caused by the formation of a more stable and cohesive solvent adlayer, impeding the access to the surface. Notably, a similar argument was proposed in a recent study, in which the capability of various peptides to bind to a metal surface was studied in water, revealing the importance of solvent structuring as an effective adsorption barrier [128].

3.3.4 Adlayer cohesion energies

To verify the hypothesis made at the end of last section, additional MD simulations were necessary, targeted to establishing the influence of surface topology on the cohesion of the first adlayer of hexadecane. To this end, a new series of systems was realised removing the dendron molecules from the starting configuration of the model described earlier, thus leaving the solvent to interact freely with the three surfaces. The calculations were then carried out as before, using the same simulation protocol and at the same temperature (300 K).

This second set of simulations confirmed that surface morphology is expected to play a decisive effect on the layering of solvent molecules. The snapshots reported in fig. 3.13, in fact, evidence

	Graphene	Graphene-$(CH_3)_{35}$	Graphene-$(CH_3)_{70}$
E_{ads} $[eV/nm^2]$	ŋ-1.18	-1.00	-0.85
E_{cohe} $[eV/nm^2]$	ŋ-0.23	-0.16	-0.11

Table 3.4: Physisorption and cohesive energies for the hexadecane adlayer: E_{surf} is the adlayer physisorption energy with the three model surfaces; E_{cohe} is the intra-layer cohesive energy.

a distinctive qualitative difference in the arrangement of the aliphatic hexadecane molecules on to the surfaces. On flat graphene, the hexadecane molecules lie in relatively straight, parallel manner, along two specific direction of the honeycomb lattice, suggesting the possibility of formation of a two-dimensional crystal on large aromatic surfaces. Importantly, the existence of a smectic phase such as this, one on both Au (111) and graphite surfaces, has been reported for alkanes of different sizes by experimental studies using in situ STM in liquids [129, 130], and it was later confirmed by some computational works [131, 132]. Consistently with our initial assumption, this ordered arrangement is progressively lost for an increasing fraction of methyl groups. Particularly, for the dense methylated surface, the aliphatic chains cannot align anymore because of the steric hindrance exerted by the high fraction of surface functionalities. These findings support our hypothesis for the delayed dendron sticking at the flat graphene surface, since a more ordered layer is expected to be more stable than a disordered one.

To investigate this point quantitatively, we computed the average interaction energies of the first layer of molecules within 7 Å from the surface (the distance corresponding to the first adlayer[1]); this quantity was decomposed into two quantities (cfr. table 3.4): E_{ads}, the interaction energy of the first adlayer of hexadecane molecules with the surface i.e. the adlayer adsorption energy; E_{cohe}, the interaction of the hexadecanes within the first adlayer with each other i.e. the adlayer cohesive energy. The energy values reported in table 3.4 quantify the observed different arrangement of hexadecane molecules at the model surfaces. The observed packing ordering on the *FLAT* graphene surface corresponds to larger adhesion and cohesive energies when compared with the other two systems. For the *FLAT* system, in fact, the aliphatic chains are able to maximise the interaction with the graphene surface while, at the same time, their parallel alignment allows for a higher cohesion. Methylation, on the contrary, causes the decrease of both the energy terms by a significant amount. Moderate methylation (*SPARSE*) led to a decreased physisorption energy, E_{ads}, of about 17%, while for the highly methylated surface (*DENSE*), the average energy contribution dropped by 38% with respect to the clean graphene system. An even higher detrimental effect was observed on the cohesive energy E_{cohe}: in this

[1]The atoms contained within the first 7 Å from the surface were selected, while for the *SPARSE* and *DENSE* systems, this distance was varied to account for an equal number of atoms.

case, the energy of -0.23 eV/nm^2 reported for the *FLAT* system was found to decrease by 64% and 136% for the *SPARSE* and *DENSE* methylated surfaces, respectively.

These results are consistent with the idea that topological defects and defective aromatic areas are responsible for a significantly reduced adhesion and a lower cohesion of the first layer of solvent molecules. In particular, the relative variation of the cohesive energy due to methylation is larger than the corresponding value for the surface adhesion energy, while both quantities have to be considered for a full understanding of the fluid/surface model. Stable and ordered adlayers of aliphatic chains, in turn, are expected to constitute a barrier for the penetration of small functionalities, hindering the interaction with the surface. Building from this topological argument, we speculate the same effect is likely to be expected for alkanes of different sizes and for different functionalisations of the surface, as independently pinpointed by Biggs *et al.* in [128].

3.4 Summary

In this study, part of the work conducted in collaboration with BP has been presented. The binding energies of several PAH dimers were calculated through DFT and compared with the values obtained for the same series of PAH molecules positioned onto a graphene plate. Importantly, the binding energies of the two series of systems showed that smaller PAHs bind more strongly to graphene than they do to each other. Thanks to this important finding, we proposed the design of a simple dendritic molecule aimed at exploiting this energy difference thanks to a unique molecular structure which combines the interactions of multiple phenyl-terminated hydrocarbon legs to stick at a graphene surface, in place of using larger moieties. We proved this to work with the aid of classical MD calculations, for which an ad hoc correction of a widely used *ff* was introduced to align the binding energy obtained with the classical potential to the DFT level of accuracy. Gas phase calculations and a set of more realistic simulations including solvent molecules and accounting for model surfaces with an increasing fraction of defective aromatic areas, have been carried out. The results supported the proposed multivalent design approach: the dendron was, in fact, observed to bind effectively to the model surfaces overcoming the competing interactions. Importantly, the sticking mechanism was observed to occur more readily on heterogeneous surfaces than on flat graphene. An explanation for this was provided by the weakening of the first hydrocarbon adlayer, a mechanism that was also observed for peptidic systems in water. Importantly, we showed that small terminal substituents are expected to adapt to the varying degree of aromaticity of the substrates, finding the most convenient adsorption sites and replacing the adsorbed aliphatic molecules.

Despite the limitations of the model, this study provided precious insight into the functionalisation of molecules to be employed in hydrocarbon mixtures. The experimental validation of the results presented in this chapter, particularly for the layering of aliphatic chains onto graphitic substrates, is currently undergoing. The functionalisation of the proposed dendron structure with different moieties targeted to a precise chemical composition of the substrate have been considered and modelled with similar simulation procedures, the results of which are not presented here, having been driven by a precise technological application, thus bearing little scientific interest.

Chapter 4

Hybridisation limit for high-density DNA monolayers

In this last chapter, we present the results of a series of MD simulations conducted on molecular systems designed to model high-density DNA monolayers, which are used by an experimental group at SISSA in Trieste to perform mechanistic studies on enzymes and for the realisation of sensors. Using a reference grafting density of oligos, we created a series of atomistic models characterised by an increasing fraction of hybridised strands, as to represent the different hybridisation states of the functionalised surface. Importantly, we show that the energy trend obtained over the range of compositions provides a rational for the hybridisation limit evidenced by independent experimental studies. We also compared the AFM height measurements with those extracted by MD trajectories, revealing the compaction effect induced by an increased concentration of monovalent salt.

4.1 Introduction

The advantages of the miniaturisation of devices has been clear since the development of experimental techniques which allowed to manipulate nano-structures, or even single atoms, with unprecedented precision. The AFM [133] (*atomic force microscop*e) and the STM [134] (*scanning tunnelling microscop*e), have contributed to the study and realisation of nano-devices, opening new possibilities in the field of electronics and for the development of technologies that today have become established realities. While at the beginning, most efforts were mainly dedicated to study of inorganic systems, materials and interfaces, biological system soon started to be in-

vestigated exploiting the newly developed techniques. The interest naturally focused on trying to bridge the gap between the usual borders of biology, biotechnology and medicine and those typically defined by solid state physics and inorganic chemistry: this multidisciplinary approach is aimed at understanding the molecular mechanisms that determine the complex metabolic network from which life originates.

In this framework, particular interest was raised by the line of research devoted to the study o fthe behaviour of biomolecules. Proteins, nucleic acids, lipids and sugars are, in fact, the main building blocks of cells: the cell membrane is formed of a double layer of lipids; its mechanical properties are governed by the formation of filaments of contractile proteins; also enzymes are formed by proteins which enable large part of the metabolic processes; nucleic acids encode the genetic information. It is clear how characterisation of the interactions between these important molecules, is of vital importance for the advancement of medicine and biology; furthermore, it also allows to exploit their spontaneous aggregation mechanisms and their chemical activity for the creation of bio-inspired technological platforms for nano-devices. Bio self-assembling showed, in fact, to be a viable bottom-up alternative to the usual top-down approach used in the fabrication of electronics. A number of works in literature have produced already sound proof for the feasibility of this concept, which pave the way for the realisation of future applications.

One of the most important and recent example of this perspective is the seminal work of Rothemund on nucleic acids [135]: he successfully demonstrated the possibility of engineering complex geometrical planar structures by combining multiple sequences of single strands of DNA (ssDNAs), able to self-assemble from solution into interlaced aggregates of the desired shape thanks only to the base paring mechanism. Besides being a demonstration of the possibilities offered by biological self-assembling, the mechanism of molecular recognition of DNA strands was also applied for the development other existing technologies, such as the functionalisation of metal surfaces for biochips and nano-arrays in bioanalytics [136].

4.1.1 DNA Self-Assembled Monolayer

A biochip is a small device commonly used in lab procedures to perform tests and analysis on molecules with a high throughput and speed. In general, a biochip is created via chemical deposition of probe molecules on a solid substrate, usually made of glass, silicon or metal. Probes are suitably designed to bind to a specific molecule present in a liquid sample (e.g. blood): once target molecules are bound to the biochip as a result of the specific interaction with the probes, subsequent methods of analysis and mechanistic studies can be carried out to measure relevant properties [137]. Similar arrays can also be realised on board of electronic circuits that are

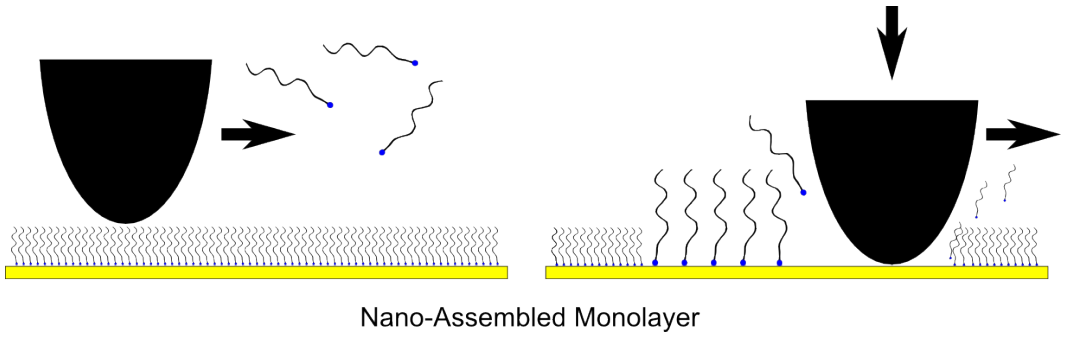


Figure 4.1: Scheme for the realisation of NAM via nano-shaving technique: a pre-existing alkyl-thiols monolayer deposited onto a metal surface is removed with the AFM tip, while the ssDNAs dissolved in the solution can bind to the surface made available by the shaving process.

capable of generating electric signals for direct output.

The most important application of this technology is adopted for the study of genetic profiles for which single strands of DNA (*ssDNA*s) can be detected according to their length and specific sequence of bases; this type of biochips have helped to dramatically accelerate the identification of the human genome and are used in forensic science for DNA tests. Other devices exploit the recognition of DNA strands for the immobilisation of proteins to study the antigen-antibody interaction and for proteomics analysis. Importantly, with the same method, it is also possible to measure the content of proteins in solution. In spite of the simplicity of the idea behind these devices, they present in practice a number of issues, such as unspecific binding events, protein denaturation, concentration thresholds and diffusion barriers [138]. For these reasons, the optimisation of the design of the bio-reagents able to target the desired molecular species in a reliable way, is a prolific field of study [139].

The realisation of monolayers can be a tricky procedure which depends on the close control of various experimental parameters. In particular, the chemical reactions which are necessary for the formation of stable bonds of probes with the surface, need to preserve the nature of the target molecule. To circumvent these drawbacks, DNA-direct immobilisation (*DDI*) [140] has progressively become one of the most common methods for creating DNA arrays. Instead of anchoring directly the selected molecule on a clean metallic surface, a matrix of ssDNAs is grafted first. In oder to do so, each strand is conjugated with a small chemical group called *linker*, which is attached to one end of each strand terminus. The linker is design to form a stable bond with the atoms of topmost layer of the (metallic) surface: as a result, the modified ssDNAs can firmly anchor to a fixed position, while the linker keeps oligos at a certain distance from the metal, thus preventing direct surface adsorption. By tuning the experimental parameters such, as the concentration of probes in the solution and the length of the DNA filaments, a variable degree

of spatial control ssDNA monolayers is achievable [141, 142, 143]. Besides, once a self-assembled monolayer (*SAM*) is formed, DNA *hybridisation* provides a simple and effective method to further functionalise the surface. Detector molecules are linked to the complementary strands, which can pair with the oligos within the monolayer, enabling a dense patterning of detectors, which is needed for the realisation of a highly sensitive device. The approach is particularly interesting because multiple different detectors can be placed into one chip for the simultaneous detection of different molecular species; besides, because of the stability of nucleic acid, DNA microarrays can be stored almost indefinitely, functionalised with the desired antibody just prior to use and subsequently regenerated by alkaline denaturation of the double helices.

4.1.2 Surface hybridisation of DNA monolayers

DNA arrays like those described in previous section are studied by a group of experimental collaborators to develop a new type of mechanical sensor capable of measuring the content of molecules in solutions up to pico-molar concentrations [144]. The strategy behind the development of the sensor is to create functionalised DNA monolayers on top of silicon pillars, in order to determine the quantity of molecules detected from the frequency shift of the pillars' normal modes of vibrations. Firstly, a self-assembled monolayer of alkyl-thiols is formed on the pillar tip: the aliphatic molecules bind to a previously deposited thin gold surface, thanks to the covalent bond formed between the sulphur atom of thiol groups and the Au topmost atoms of the surface, while a dense monolayer is achieved as a consequence of the favourable dispersive energy resulting from inter-chain packing [145]. Subsequently, DNA nano-patches are realised within the alkyl-thiol monolayer thanks to a technique called *nano-shaving* [146]. By applying a force between 10 and 100 nN with the tip of an AFM, the grafted molecules are locally removed with precision (see fig. 4.1). At the same time, the ssDNAs in the solution can bind to the exposed surface thus replacing the alkyl-thiol molecules. With this procedure, the density of DNA monolayers can be closely controlled upon varying the shaving parameters, i.e. the applied force and the scanning tip speed [147]. Experimental evidences showed that the order and the density of the *nano-assembled monolayers* (*NAM*) are significantly increased with respect to a spontaneous growth of monolayers from a concentrated solutions of precursors: ssDNAs can, in fact, bind directly in an upright position, allowing to reach higher densities [148].

As for SAMs, NAMs can be functionalised via hybridisation of the monolayer with the complementary DNA strands. When a fluid sample is placed above the sensor, a certain amount of target molecules binds to the monolayers. This has the effect of varying the mass distribution of the pillar, thus affecting the frequencies of the normal modes of vibration which can be

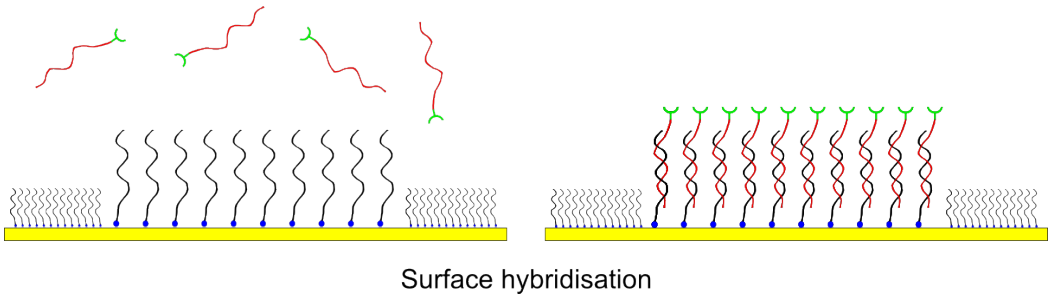


Figure 4.2: The DNA monolayers realised via nano-shaving technique can be conveniently functionalised via the hybridisation of complementary strands exploiting the base pairing mechanism. Each complementary strand is conjugated with a detector molecule (rendered in green), which is designed to bind to the molecule of interest.

measured with optical techniques. By modelling the kinetics of diffusion process toward the surfaces, the frequency shift can be related to the concentration of targets molecules in the solution. Importantly, it has been demonstrated that the pillars not only enable the frequency measurement with precision, but allow to decrease significantly the diffusion time, thus increasing the sensitivity of the device [144] when compared with sensor built on planar surfaces. This fact is extremely important for applications in medicine and biology, for which similar mechanical biosensors demonstrated to provide reliable measurements and short time responses [149].

The feasibility of this idea of building nano-mechanical sensor based on silicon nano-pillars has been already demonstrated in practice [144]. However, to optimise the design of the device to increase performances and ease fabrication, it is crucial to study and control the experimental parameters. The first problem to face in this framework concerns the kinetics of the diffusion process: in the limit of low concentration, the transport of targets to the surface is the limiting phenomenon, in particular when the molecules are relatively large proteins and signal amplification strategies, such as polymerase chain reaction, are not desired or not viable. In these cases, the binding rate is often controlled by diffusion: short saturation times (in the range of minutes) guarantee fast measurements and feasible experimental procedures. Another complicated issue is constituted by *surface hybridisation* which, as outlined earlier, is the key mechanism used to functionalise the DNA monolayers. Hybridisation occurs spontaneously in water at ambient condition; however, when DNA strands are tethered to a surface the process is complicated by the increase charge concentration and not yet fully understood [150, 151]. In this framework, theoretical models and computer simulations can provide a substantial contribution to ease the interpretation of experimental results.

4.1.3 Hybridisation limit of high-density DNA monolayers

Many experimental works have addressed the effect of the grafted probe molecules density on the hybridisation efficiency, i.e. the fraction of grafted strands successfully hybridised within the surface. For the case of short DNA chains, the typical coverage values in the saturation limit, vary from 1×10^{12} up to 3×10^{13} $strands/cm^2$ [144, 152]. The upper density value was observed for the grafting of ssDNAs, which are characterised by a lower charge density, a smaller diameter (~ 1 nm); the lower density value, instead, was measured for the grafting double stranded DNA strands ($dsDNAs$), which were subsequently denatured to remove the complementary strand. For lower coverages, the totality of the grafted strands can be hybridised because of the wide separation between adjacent chains, resulting in a low electrostatic repulsion; this corresponds to a hybridisation efficiency of 100%. Differently, for a higher coverage, in the range of $1 - 2 \times 10^{13}$ $strands/cm^2$, a decreased hybridisation efficiency was reported by many authors. Georgadis et al. [152], for instance, reported efficiency of about 10% for spontaneously grown SAM. This result was shown to agree with the statistical models implemented by Chakraborty et al. [153] for grafted oligos. Nano-ordered arrays obtained via nano-shaving technique, on the other hand, showed to yield higher hybridisation efficiencies and, in particular, for the same coverage of grafted probes. For a probe density of about $1 - 3 \times 10^{13}$ $strands/cm^2$, for example, values of 30% have been independently reported by several authors [144, 147, 154]. The consistency of these findings pertaining to independent works, suggests the existence of an hybridisation limit for DNA *high-density monolayers* (HDM).

Several different factors have to be taken into account for the characterisation of DNA monolayers. The first important characteristic is the large charge density of nucleic acids which is known to greatly influence their biological and chemical behaviour by determining DNA packing, pairing, and the interactions with other macromolecules [155, 151, 156]. In DNA monolayers, the combination of negatively charge filaments and the spatial confinement, greatly enhance the electrostatic repulsion between DNA strands. In addition, DNA electrostatics is affected by the conterions present in the solution, which screen the DNA charge on the scale of the Debye length. Charge screening, given by the interplay between solute-ion electrostatic attraction and thermal motion, plays a fundamental role in counterions condensation, a characteristic phenomenon observed in charged polymers as elucidated by Manning [157]. Electrostatic repulsion between oligos and ion screening have been shown to significantly influence the properties of DNA monolayers. AFM imaging techniques, in fact, showed that the height of SAM is proportional to the density of grafted molecules and how this dependance is sensibly reduced for an increased ionic concentration [147, 158]. AFM measurements have also been used to monitor the occurrence of hybridisation in monolayers. The persistence length of ssDNAs is, in fact, almost

one order of magnitude lower than that of dsDNAs [159, 160]: as a result, monolayers formed by single and double strands of DNA behave quite distinctively when a mechanical load is applied by the AFM tip, thus the occurrence of hybridisation can be easily detected. This principle, was further exploited to relate the mechanical properties of monolayers with their composition, density and hybridisation state [161].

The effect of steric hindrance induced by molecular crowding is another important element controlling the properties of DNA monolayers. Given a certain coverage, in fact, the spatial ordering of the molecules in the monolayer is almost as important as the probe density itself. This has been highlighted by a series of experiments where the influence of the arrangement of molecules on the functionalisation capability of a DNA array was assessed by measuring the occurrence of a restriction reaction mediated by enzyme of different size: this experiment showed that enzymes (and similarly ssDNAs), are adsorbed into the monolayer from the solution, but that a necessary step for hybridisation and the restriction reaction is the lateral diffusion of adsorbates, which was found to be partially hindered for partially matched DNA sequences [162]. Here, the role of the charge screening is again fundamental, since it was shown affecting both the grafting density and hybridisation. In this framework, however, it is not clear how ions might influence the conformations of DNA strands and the resulting packing structures [163], nor it is the effect on the hybridisation binding energy.

4.1.4 Computational contributions

The study of DNA monolayer has already been tackled by a number of heterogeneous works in which different computational techniques have been applied to elucidate some of the aspects described earlier. A first group of works is based on the use of lattice models that treated grafting and surface hybridisation with a stochastic approach. Schatz et al. [148], for instance, tried to the calculate the binding rate coefficients by modelling the DNA grafting with a first order Langmuir reaction and compared the Monte Carlo simulations results with experimental deposition data. A similar approach was followed by Chakraborty et al. [153], who studied the surface hybridisation of grafted DNA with the *Poland-Scheraga* model [164]. Analogously, Monte Carlo simulations employing classical potentials have been used to study the effect of probe density and probe length on hybridisation [165]. According to these works, surface hybridization of DNA HDMs is expected to occur in two consecutive steps. Firstly, the complementary oligos are adsorbed in the monolayer from the solution: the adsorption expected to saturate early with the oligos' concentration and is followed by the diffusion of the strands within the array. Strand pairing was shown to be a relatively fast process for short strands. However, the limiting step

of hybridisation was reported to be the nucleation of few metastable base-pairing interactions, followed by zippering of the remaining bonds of complementary oligos [166, 167]. The picture emerging from all these observations indicates that the intrinsic hybridization limit of HDMs is tightly related to the structural features of the nanostructure.

In this context, the intrinsic hybridization limits of DNA HDMs can be tentatively rationalised by the fact that both charge and atomic density increase with hybridization inside the patch. In 2007, Yao et al. [168] highlighted some results suggesting that the equilibrium distance between thiolated dsDNA molecules grafted on Au (111) is higher than that between ssDNA. Despite the computed equilibrium densities did not compare well with the available experimental values, possibly due to technical limitations of the approach adopted, their work suggested qualitatively that an energy penalty is be associated with DNA molecules too tightly packed inside the patch. This would appear reasonable, since electrostatic repulsion generally increases when DNA molecules are arranged tightly closer. This also agrees with the findings of Pinheiro et al. [169], who demonstrated how hybridisation rates are influenced by the presence in neighbouring sites of dsDNAs instead of ssDNAs.

Atomistic simulation studies have also been used to provide insight into HDMs. For their importance as precursors for the realisation of biochips, the formation of alkyl-thiol monolayers have been studied by classical MD [170] and coarse-grained models [171, 145] to elucidate the growth mechanism of dense arrays from an initial sparse disordered phase, and the resulting structural properties. However, since size of DNA oligos and the timescale during which hybridisation is expected to occur is out of reach for most atomistic models, DNA denaturation was investigated instead. The reaction pathway leading to strand separation has been studied by means of Steered Dynamics [172] by imposing an asymmetric helix stretching, evidencing some aspect of the unwinding of double helices. Other studies focussed on the description of conformational properties of DNA duplexes bound to a gold surface. Notably, Lee et al. carried out a series of studies to determine the effect of salt concentration on the stabilisation of packed oligos bound to a surface [173], for the functionalisation of gold nano-particles [174] and in the base pairing of complementary sequences for nano-particles self-assembling [**?**]. Although these recent works made important steps towards the characterisation of HDM, a comprehensive understanding of the factors that control hybridization in these high-density nano-structures is still missing. Here, we aim at filling this gap, by providing insight in the molecular origin for the defective strand pairing in HDMs.

Figure 4.3: DsDNA sequence used for the HDM models and for the AFM experiments. The 5' terminal nucleotide of one of the two strands is covalently conjugated to a hexane thiol: the lone pair of electron of the sulphur atom is able to form a covalent bond with the gold atoms of the surface. For ssDNA only the top sequence was used.

4.2 HDM model and computational details

4.2.1 Modelling the DNA monolayer

In order to elucidate the factors that control and limit the hybridisation of DNA monolayers, we developed an atomistic model aimed at evaluating the energetics of DNA HDM and the effect of the increased molecular crowding that is reached upon hybridisation of a realistic DNA patch. This model was also conceived to complement the ongoing AFM characterisation of analogous systems carried out by a experimental group at SISSA, whose measurements will be used here for comparison. The atomistic approach presented herein, allowed a detailed molecular-scale investigation of the HDM hybridisation limits, including an explicit representation of the solvent and the electrostatic interactions occurring in the system.

Clearly, a full account of the complexity of a monolayer of micrometric size is out of reach for MD simulation with full atomistic detail. A reduced representation is, therefore, required to represent an extensive system in a computationally feasible way, yet conserving the essential physical properties of the real system. In this context, periodic boundary conditions can be conveniently exploited to limit the study of a small portion of a monolayer that is representative of a real, more extensive HDM. Unfortunately, no detailed information of the structure of a monolayer as such has ever been made available in literature. However, the densities of grafted strands can be extrapolated by AFM measurement; moreover, we know from other studies [143, 163] that the packing order of nucleic acids onto a surface (for a sufficiently high density) is expected to be hexagonal. These two elements together allowed us to build the base model system of this study.

The reference target density of grafted DNA strands used in the experiments corresponded to $1.2 \times 10^{13} strands/cm^2$ and was used here. The thiolated DNA strands of base sequence shown in fig. 4.3, were arranged in 3 x 3 hexagonal close-packed array: the nearest neighbour distance between thiolated DNA molecules was set to 3.1 nm, which allow to match the experimental density value. The DNA packing structure was then placed on top of a rhombic 5 layer Au (111)

slab of dimension 9.7 x 9.7 x 1.1 *nm*, with the alkyl-thiol linker placed at distance of 2 Å from the topmost Au layer, roughly corresponding to the length of the bond Au-S bond. In fig. 4.4, a rendering of system is displayed, showing the initial set up of a reference HDM model system formed exclusively by ssDNAs, being representative of a DNA HDM before hybridisation.

The strategy adopted in this study for the representation of different possible HDM hybridisation states is summarised in fig. 4.5: a total number of ten HDM systems, corresponding to as many different hybridisation fractions, were created by substituting the ssDNA with a variable number of dsDNA molecules. In fig. 4.5, the systems are labelled with their respective hybridisation fractions, from HDM 0%, where all grafted oligos are ssDNAs, to HDM 100%, composed by just dsDNAs. This set of systems aimed at identifying the different favourable and unfavourable factors which determine the hybridisation fraction limit for a given density DNA HDM. This is consistent with a physical picture of HDMs formed by thiolated ssDNA immobilised at the surface where hybridisation is an initially highly favoured process, which gets progressively hindered by increasing charge and atomic density. For this reason, since the hybridisation of additional oligos is expected to be accompanied by progressively unfavourable energy contributions, nearest neighbour dsDNA molecules were avoided in the initial, low hybridisation fraction systems. This is consistent with the relatively large optimal distances between dsDNA predicted by Yao et al. [168], and, more generally, with the expectation that, once a ssDNA is hybridised, the enhanced charge and atomic density around the newly formed grafted dsDNA hinder the hybridisation of the adjacent ssDNAs. This is also consistent with the results reported recently by Pinheiro et al. [169], who have shown that the hybridisation rate is higher for hybridisation sites located in the outer, rather than inner regions of a grafted DNA tile, or when the nearest neighbour sites host ssDNAs instead of dsDNAs [169]. As shown in the scheme of fig. 4.5, however, the presence of neighbouring dsDNAs can be completely prevented only up to the HDM 33% system

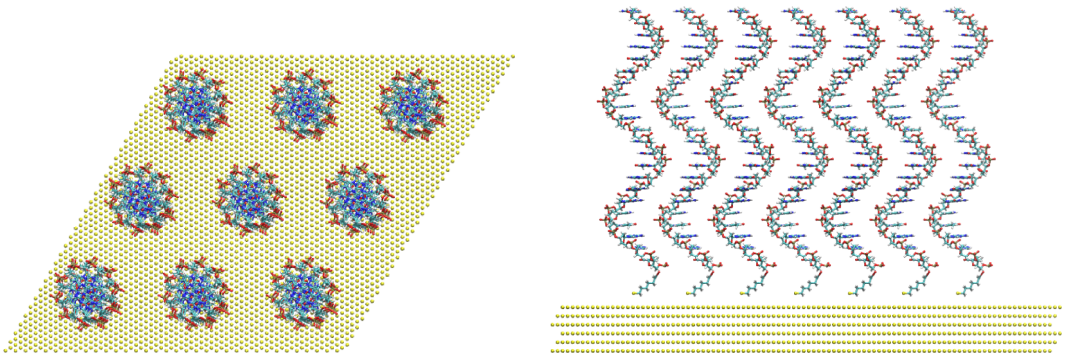


Figure 4.4: Starting configuration of the HDM 0% model system: a 3x3 array of thiolated ssDNA strands is placed above the rhombic Au (111) slab; this system represents a HDM of hybridisation fraction equal to 0%.

(3 dsDNAs for 33% hybridisation).

The simulation box of each grafted HDM system was filled with TIP3P explicit water molecules and Na+ and Cl– ions to neutralise the net charge of the system (required by the convergence of long range electrostatics) and to reproduce the experimental ionic strength. The number of ions was calculated from the equilibrated volume by subtracting from the total volume occupied by solute molecules. The target concentration of monovalent NaCl was set for the first set of simulation to 300 mM, a value in the range of those commonly used experimentally to perform AFM measurements; additional two salt concentration of 200 mM and 400 mM were used for the comparison with experimental results.

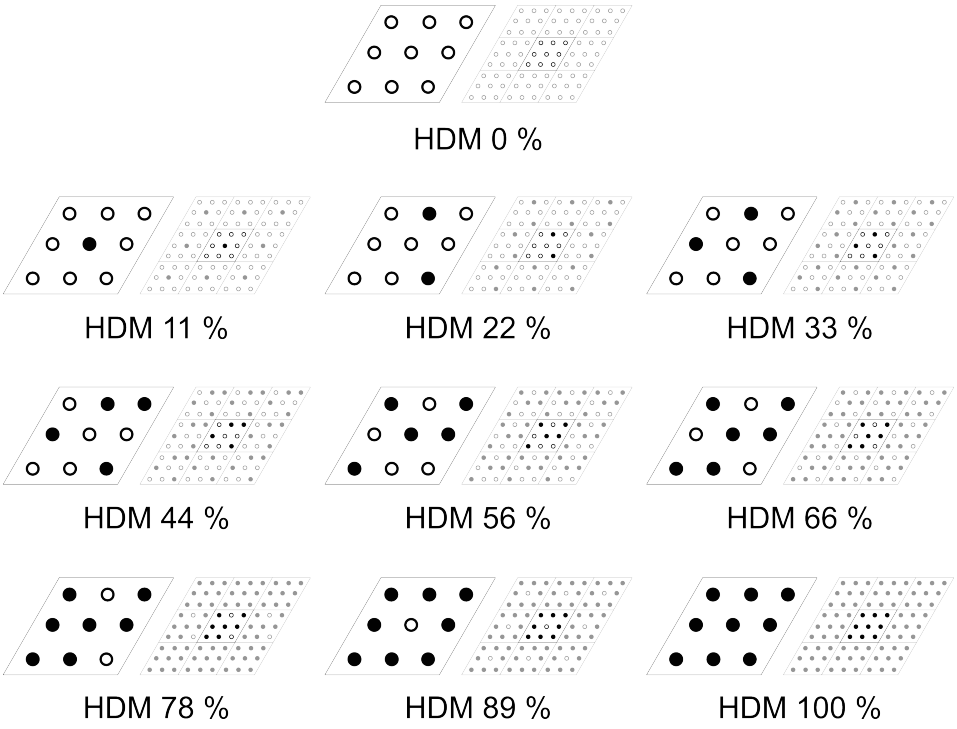


Figure 4.5: HDMs models for variable hybridization fraction: the grafting of 9 ssDNA strands (empty circles) represents the DNA grafted monolayer before hybridisation (HDM 0%); an increasing number of ssDNAs is substituted with dsDNAs (full circles) as to form monolayers of different composition and hybridisation fraction, from HDM 11% to HDM 100%, the latter representing a completely hybridised patch.

4.2.2 Parametrisation and simulation procedure

The set up of the model is not complete without a proper choice of a *ff* that accurately represents the structural feature and the interaction of nucleic acids. As described earlier, particularly important is the account of the electrostatic interactions which are fundamental for confined

system of nucleic acids systems, such as DNA monolayers; the peculiar charge distribution of the phosphatic group was shown, in fact, to be fundamental in the packing of long DNA strands in viral capsides [175].

A conspicuous number of *ff* has been proposed over the years, which have been constantly improved to match both experimental's and first principle calculations' results. The *AMBER ffX* series groups successive refinements of the original implementation of the *ff* for nucleic acids developed by Kollman et al. [176], which showed to reproduce the structural features of a large variety of nuclei acids in water [177]. A Recently development based on on high-level QM calculations for the tuning of torsional angles, led to the release of the ff99bsc0 *ff* [178] which is, at present, the most accurate potential available for B-DNA and RNA and was, therefore, chosen for this study.

The grafting of the DNA sequences was made possible by the formation of a strong covalent bond between a gold atom of the metal surface and the sulphur atom of the alkyl-thiol linker, which is chemically linked to the 5'-end terminal base of each DNA strand (ssDNA or dsDNA). A 6 carbon aliphatic chain was used for this system, analogously to the experimental set up (see below), and it was parametrised separately using the *GAFF ff* [14]; eventual missing parameters and the partial charges were calculated with the semi-empirical method AM1-BCC [120]. The coordination bond between the thiol groups and the surface gold atoms, was modelled through a harmonic potential, consistently to what done in other studies: the equilibrium bond length and force constant were taken from [170]; these values were derived from ab-initio calculations and have been tested for the interaction of alkyl-thiol monolayer onto an Au (111) surface. Since this study focused on the energy of HDM only, we did not intend to model explicitly the motion of metal slab's atoms: on the contrary, the gold atoms were kept fixed at their initial position with strong harmonic positional restraints. The non-bonding interactions of the Au atoms with the rest of the atomic species of DNA and water molecules were set using a 12-6 LJ potential, using of the parametrisation (σ=2.569 Å and ϵ=0.039 *kcal/mol)* adopted in [173].

The starting configurations were generated with the aid of Python routines which were written for building the Au slab, placing the DNA chains as displayed in fig. 4.4, adding water molecules and ions. The simulation procedure followed for carrying out the dynamics of the HDMs, consisted of few preliminary step for (i) heating the system to the target temperature (300 K), (ii) equilibrating the density and the size of the simulation box, (iii) equilibrating of the DNA strands and, finally, (iii) the generation of the equilibrated trajectories for the thermodynamic averages. The details of each step of this protocol are summarised as follows:

- the minimisation of the initial structures has been performed to remove unwanted contacts

between solute and solvent atoms; the steepest descent method has used for the first thousand iterations, followed by a conjugate gradient algorithm till convergence; restraints have been applied to the solute (DNAs and gold atoms).

- the system has been heated up from 0 K to 300 K with a progressive increase of temperature occurring in a total simulation time of 200 ps; restraints of solute atoms have also been applied during these first steps of dynamics;
- the equilibration of the solvent was achieved within the following 7 ns of simulations; this time was required to equilibrate the water to its bulk density and allowing the diffusion of ions with in the rhombic simulation box; the cell dimension in the direction orthogonal to the gold surface has been adjusted repeatedly to account for the density variation, since the available NPT algorithm did not allow to decouple pressure control along different axis; the final box height was about 150 Å, allowing enough space between replicas.
- upon density equilibration, the proper MD run has been executed for a total simulation time of 100 ns, the last half of which has been used for the analysis.

The dynamics has been performed in NVT ensemble with a Langevin thermostat (collision frequency equal to 2 ps^{-1}) at the temperature of 300 K. The leap-frog integrator has been used with a time-step of 1 fs during the first simulation stages (heating and density equilibration) while during the equilibration dynamics the time step has been set to 2 fs. The cut off radius for short-range interaction has been set to 1.0 nm while the long range electrostatic contributions have been treated with the Particle Mesh Ewald method (PME). The $SHAKE$ algorithm [179] has been applied to bonds and angles involving hydrogens. The total length of the simulations has been divided into runs of 10 ns each. The final trajectory has been generated saving a system configuration each 200 ps, collecting 500 frames in total. The equilibrium energy averages ($\triangle E$) for the HDM systems were carried out with the MM-PBSA approach [44] (see sec 1.3.2), according to which the energy of each system is derived as the sum of solute-solute and solute-solvent interactions through a continuum solvation model. All MD simulations were carried out using the AMBER 11 MD package, compiled to run on a GPU graphic card.

4.2.3 Experimental setup

Real DNA monolayers were realised in laboratory at SISSA. Ultra-flat gold samples were first immersed in a solution of thiolated oligo-ethylene-glycol ($TOEG$) at 4 C for 12-18 hours. The SAM formed was then rinsed in ethanol, dried and glued in a AFM liquid cell. The nano-grafted DNA HDM was subsequently obtained by applying a load of about 80-100 nN with

the AFM tip on a selected area of the monolayer in the presence of thiolated ssDNA in a 1 M solution of NaCl; the high load applied promoted the replacement of the TOEG molecule with the thiolated DNAs. By resetting the force exerted by the AFM tip to a lower value (∼0.1 nN), the DNA monolayer was imaged and the height accurately measured. The hybridisation reactions were carried on at room temperature within the AFM liquid cell in using a highly concentrated solution of the complementary sequence. The AFM height measurement before (HDM 0 %) and after (HDM 33%) hybridisation were performed in non contact mode at different salt concentrations (200mM NaCl and 400mM NaCl) with standard silicon cantilevers to avoid compression of the DNA layer.

4.3 Energy and structural analysis

4.3.1 HDM energy profiles

The energy for each system was averaged over the trajectories and calculated as the sum of the potential energy of DNA strands and the solvation energy term, for which a continuum model was employed (replacing explicit water molecules). This approach was chosen to cope with the different number of water molecules used for the series of HDM systems[1], which did not allow to compare directly the total potential energy. Using a continuum model for solvation based on the solution of the Poisson-Boltzmann equation, instead, it was possible to substitute the contribution of explicit solvent molecule with an implicit term calculated from solutes' trajectories, as described in sec. 1.3.2. Additionally, to compare the results accounting for a variable number of DNA strands, the system energy was subtracted by that one of the HDM 0% system, used as a reference, and normalised by the number of strands contained in each system.

The energy differences $\triangle E$ are reported in red in the left panel of fig. 4.6 as a function of the hybridisation state. This energy profile identifies energetically favored or disfavored hybridization states with respect to the HDM 0% system, which represent the DNA monolayer before hybridisation. The plot shows that the energy decreases for hybridisation fractions up to 33%, evidencing that the pairing of additional oligos is favoured. The curve displays a minimum for the HDM 33% system, for which three out of nine strands within the monolayer are paired: a corresponding energy difference of about -16.9 ± 5.0 $kcal/mol$ per strand was calculated respect to the un-hybridised system. The energy plot also shows that only hybridization fractions lower

[1] Although the volume of the cell was kept constant for all simulations, the number of ions needed to keep a fixed concentration depends on the actual water volume, which varies with the number of dsDNAs; the water volume was calculated as the equivalent volume occupied in bulk at the same temperature by the same number of water molecules. The minimum value of the cell volume required for the given ions concentration, was therefore determined upon respecting the charge neutrality constraint.

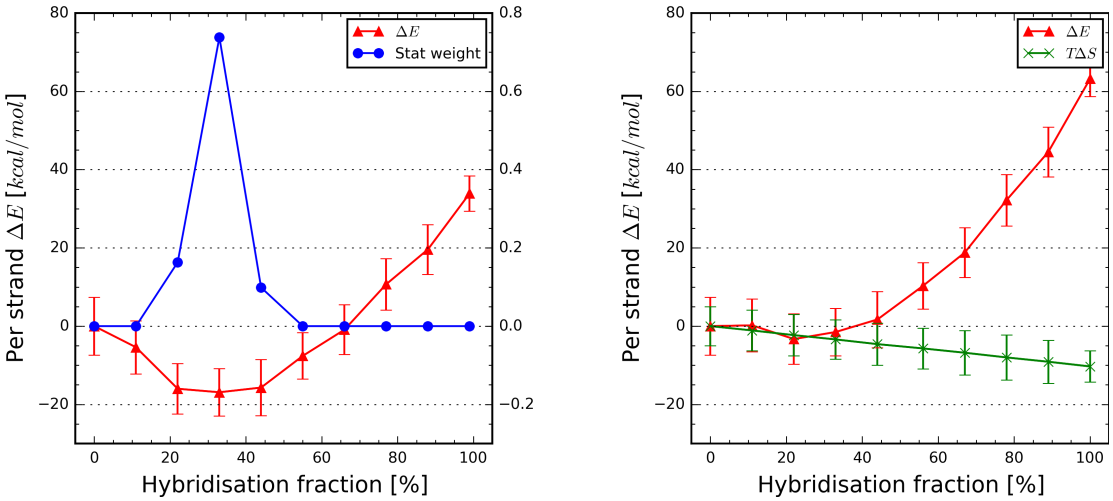


Figure 4.6: Energy differences calculated for the systems of increasing hybridisation fractions with respect to the un-hybridised system (HDM00%), used as a reference. *Left*: the average energy for each system is normalised over the number of oligos: ΔE values identify energetically favored or disfavored hybridisation fraction. The statistical weight given by the calculated energy difference is reported in blue. *Right*: the energy of the HDM systems for a constant number of strands was obtained by adding the contribution of additional strands in solution, calculated separately. The energy differences with respect to the un-hybridised system do not vary significantly for hybridisation fractions below 33% while increasing positive values of $\triangle E$ evidence that hybridisation fractions above 33% are progressively unfavoured. In green the entropic contribution was estimated via normal mode analisys.

than 44% are, in practice, energetically allowed for a DNA HDM of the present density. The hybridisation of additional ssDNAs above the 33 % threshold, in fact, was found to result in an energy penalty, which reverted the energy profile: above this limit, indeed, strand pairing becomes unfavourable and the system appear to have reached its crowding limit. This effect is limited for the HDM 44% system, for which an increase as little as 1.2 $kcal/mol$ per strand was determined. However, for higher fractions, the energy increase is more pronounced. In particular, for hybridizations higher than 67% (HDM 67%) positive values of $\triangle E$ suggest that the favorable strand pairing is likely hindered by unfavourable electrostatic and van der Waals repulsions between adjacent strands. It is worth noting that this approach, by not considering explicitly the diffusion and the adsorption of the complementary oligos inside the monolayers (thus neglecting the kinetic of the surface hybridisation process), only takes into account the initial non-hybridized systems (HDM 0%) and different possible final HDM states for variable hybridization fractions. As a consequence, these results pertain to a hypothetical infinite reaction time, and express the effect of increasing molecular and electrostatic crowding that follows hybridization.

The plot labelled in green in the left panel of fig. 4.6 represents the statistical weight for the

different hybridized HDM systems, which was simply estimated by the Arrhenius formula for thermally activated process as $\exp(-\triangle E/kT)/Z$. This plot provides qualitative information on the probability of different hybridization states, evidencing how the energy minimum found in precedence at 33% corresponds, at the simulation temperature of 300 K, to a significant likelihood. The same does not hold for all other states for which, with the exception of HDM 22% and HDM 44% systems, the statistical weight is essentially null: the corresponding fractions are not, therefore, expected to be likely found within monolayer of analogous densities. This finding agrees well with the experimental observations, which indicated a 30–40% maximum hybridization efficiency for a 22 base-pair nano-grafted DNA HDMs. Moreover, the energy plots substantiate the evidences reported in literature for the hybridisation limit of HDMs.

The series of HDM systems was conceived as having a variable number of DNA strands, increasing from 9 to 18 for the HDM 0% and HDM 100%, respectively. This fact was taken into account by normalising the energy of each system over number of strands. However, to validate the results presented earlier and to check whether the different number of oligos considered could have affected the energy profile, the contribution of an adequate number of unbound ssDNAs needs to be taken into account in order to compare data for a constant system composition. To this end, an additional system was created in order to determine the energy of a ssDNA of the same base sequence in solution. An additional 22 base pair strand was placed in a conveniently large simulation box (6.0 x 6.0 x 12.0 *nm)* containing explicit water molecules and counterions to reproduce the same condition of the HMD systems (∼300 mM NaCl). As before, the system was equilibrated for 50 ns at ambient pressure, while the following 100 ns (in the NVT ensemble) were used for thermodynamics averaging, according to the same procedure described earlier. The resulting energy multiplied by the number of oligos needed to match the total number of strands, was then added to the previously calculated data points.

The right panel of fig. 4.6 shows the energy differences with respect to the HDM 0%, again normalised over the number of strands, which was, in this case, constant (18). The contribution of the complementary strands has the effect of altering the shape of the energy profile without substantially affecting the conclusion drawn in precedence. The plot, in fact, evidence a negligible energy difference for hybridisation fraction lower than 44%: in practice, for the extended system formed by the HDM and the free strands, the strand pairing is counterbalanced by the increased the strand repulsion of up to 4 dsDNA out of 9 sites. Above this fraction, an increasing positive energy difference was observed, meaning that additional hybridisation is overall unfavoured: the remaining complementary strands are therefore expected to remain dispersed in solution.

These results identify an explicit energy criterion for the hybridization limit of DNA HDM which rationalises the experimental evidences reported by several authors. The high-density packing

of negatively charged oligos strands grafted onto the Au surface, despite the presence of a large number of screening ions, results in a positive energy difference that prevents hybridisation to occur above a certain threshold. This effect agrees with the available experimental evidences and can interpreted as the consequence of an increased electrostatic repulsion, which controls the maximum HDM hybridization limit. This, in turn, can further explain the packing-induced denaturation of dsDNA monolayers reported to take place onto curved [180] and flat [181] Au surfaces.

4.3.2 HDM Structural analysis

To further validate this study, we turned to the investigation of the HDM heights with respect to the Au surface, a quantity that can be accurately monitored by AFM experiments. The heights of the monolayers were measured with the AFM before and after hybridisation for two concentrations of the NaCl buffer (200 and 400 mM). The data reported in table 4.1 shows the measured height profiles and the reconstructed micrographies of the surface. The DNA monolayers are rendered by square brown patches which were created via the nano-shaving techniques within previously grown TOEG monolayers: the quantity measured is the vertical displacement of the AFM tip with respect to a reference quote i.e. the height of TOEG monolayers, from which the HDM's heights could be obtained, once the level of the metal surface was known.

The average data gathered in table 4.1 shows that the heights of partially hybridised patches are greater than for un-hybridised ones at the tested salt concentrations. This is consistent with the persistence lengths of dsDNA and ssDNA, which are known to differ by more than one order of magnitude [147]: short dsDNA behaves, in fact, as rigid rods and maintain a helix structure that prevents them from folding toward the surface, while ssDNA are expected to adapt more easily to the available space, yielding a denser monolayer. Furthermore, the ion concentration is observed to lower the DNA patch height: this effect, as previously posited in the framework of polymer brush theory [182, 183], is expected to be caused by the charge screening exerted by mobile ions on to the negatively charge DNA strands, thus decreasing the repulsion and enabling a denser packing of the oligos.

Four additional molecular systems, corresponding to the two hybridisation fraction of 0% and 33% were investigated with MD for comparison. The systems aimed at representing the DNA monolayer before (HDM 0%) and after hybridisation (HDM 33%): in agreement with the findings reported in previous sections, 33% is the expected hybridisation state of the monolayer upon reaction with the complementary oligos and it was used here as reference. Each system was simulated at the two ionic concentration used in the experiments, by adding a variable number

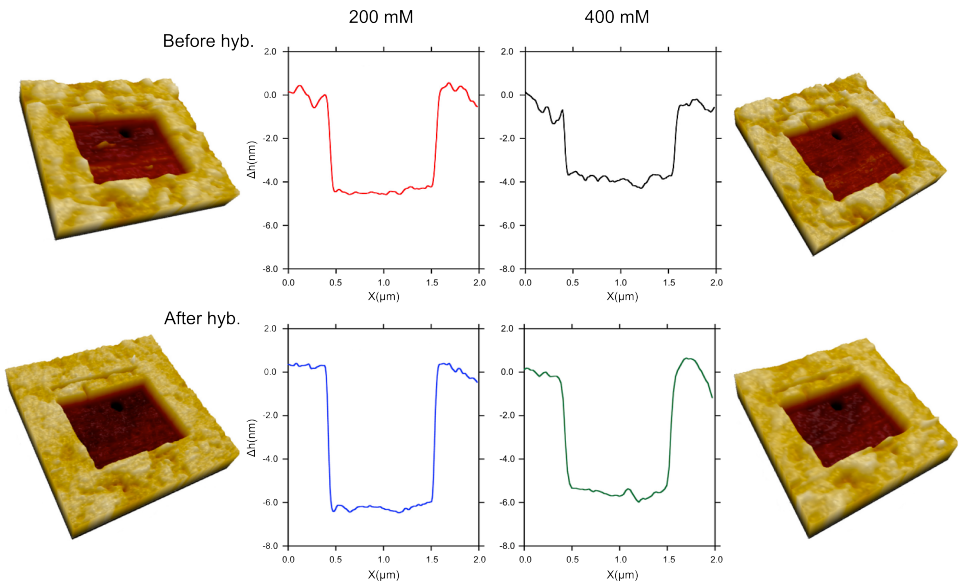


Figure 4.7: AFM experimental height measurements of two HDM systems before and after hybridisation conducted at two ion concentrations of 200 mM and 400 mM. The height profiles show the AFM tip's vertical displacement across the DNA HDM patches with respect to the pre-existent *TOEG* monolayer level; the renderings on the sides show a reconstruction of the measured HDM surface: dark squares identify the DNA patches, while the surrounding areas, rendered in yellow, pertain to the *TOEG* monolayers (courtesy of Casalis et al. [184]).

of ions; the procedure described in sec. 4.2.2 was again followed for the dynamics, while again the last 50 ns of the trajectory where used for analysis. The HDM heights were estimated through the analysis of the equilibrated trajectories; this was achieved by averaging the distance from the center of mass of the topmost residue of each oligos and the Au surface over the trajectories for each system. The results were then compared with those measured with AFM.

	200 mM		400 mM	
	AFM	MD	AFM	MD
Before hyb./HDM 0%	6.34 ± 0.40 nm	7.7 ± 1.3 nm	6.01 ± 0.63 nm	6.8 ± 0.9 nm
After hyb./HDM 33%	8.45 ± 0.56 nm	8.5 ± 1.1 nm	7.53 ± 0.34 nm	7.6 ± 0.9 nm

Table 4.1: Comparison of the heights for the DNA monolayers before and after hybridisation measured by AFM at the two ion concentrations and the corresponding data extracted from the MD trajectories for the HDM 0% and HDM 33% systems.

The data reported in table. 4.1 reveals that, consistently with what observed before, a higher hybridisation fraction and a higher salt concentration determine a larger height of the DNA monolayers. The comparison of the individual values is particularly good for the HDM 33% systems (second raw of table 4.1): a difference of just 0.1 nm was, in fact, observed with respect to the AFM data at both concentrations. In the case of the un-hybridised model system HDM 0%, the computed estimated heights are larger than the measured AFM values (first raw of table 4.1): a difference of 0.8 nm was determined for the average heights of the system at 400nM, while, at the lower salt concentration, a larger difference was observed, equal to about 1.4 nm. The offset of the average data is counterbalanced by a large error, which allows an overlap of the confidence intervals. Interestingly, the uncertainty of the measurements was found to depend significantly on the ions concentration.

The results outlined above concerning the different agreement found for the HDM 0% and the HDM 33% systems can be interpreted as an effect of the different mechanical rigidity of ssDNA and dsDNA. Non-contact AFM measurements are intended to avoid any compression of the patch; however, a residual force is necessarily exerted by the AFM tip, estimated to be in the order of 50–100 pN: this force is expected to have a different effect on HDM 0% than on the more rigid HDM 33% systems. The heights predicted by our computational model are, in fact, larger than for the measured AFM values for the very flexible HDM 0%, since the simulations were conducted at zero force. A similar consideration can be applied to the higher salt sensitivity predicted by MD for HDM 0% respect to the AFM case. A reduced saline content is, in fact, expected to screen less efficiently the repulsion between ssDNA strands, which are then hindered from forming a denser, thinner monolayer; on the contrary, higher salt concentration favours the formation of a more compact arrangement of oligos. In this respect the effect of the residual compression on the measured heights for the ssDNA system (HDM00) is expected to be larger for the lower NaCl concentrations (where we estimate it to yield an 18% height contraction) than for higher concentration (where it reduces to an 11% height contraction).

These observations are supported by the analysis of the individual strands' height, where the effect of differing NaCl concentration is more evident. In the top row of fig. 4.8, the heights of HDM 0% system shows how ssDNAs are highly responsive to an increased ion screening (top right): a significant decrease in the patch height is, in fact, enabled by the enhanced folding of several oligos toward the surface. On the other hand, the greater rigidity of the double helices is evidenced by the analogous plots for the HDM 33% systems (bottom row of fig. 4.8). The dsDNAs displayed overall a constant height for both systems: folding and bending, in fact, are essentially not possible for short dsDNAs without the unwinding of the double helix, while their rigid motions and small rotations are still allowed by the aliphatic linker.

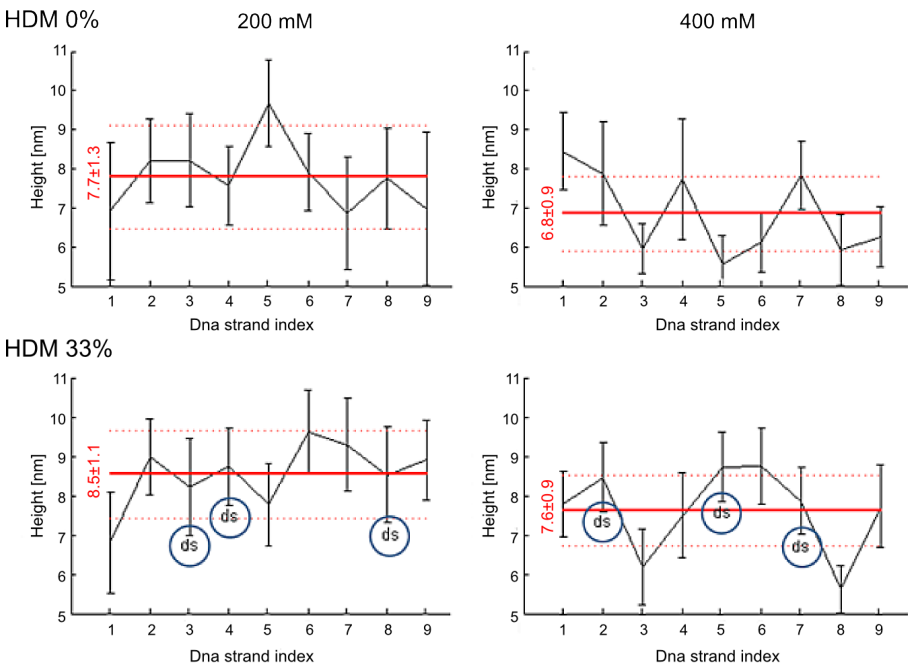


Figure 4.8: Average heights of each individual strands for the four systems: the red solid lines represent the average value and the red dashed lines define a confidence interval. The height of dsDNAs is indicated with blue circles.

The atomic density profile, $\rho(z)$, of the grafted molecules in the direction orthogonal to the surface (z axis) was extracted from the equilibrated MD trajectories for the four systems. This quantity is the central in classical DFT methods, which were developed in the context of polymer brush theory, and it is known to characterise the packing density of grafted polymers [182, 185]. The calculated profiles (fig. 4.9) are consistent with those found in studies of similar systems and bear information on the distribution of DNA atoms at different distances from the Au (111) surface, showing how this depends on hybridization and ionic strength. The density of each monolayer rapidly increases from zero within the first nm, since the oligos are kept at certain distance from the metal surface by the aliphatic linker; a plateau is reached soon after, reflecting the different packing states reached in the HDMs. Some density fluctuations are observed near the grafting surface: these oscillations are caused by the increased spatial correlation of the first bases and are proved to particularly enhanced for dsDNAs [185]. The density values in the plateau are inversely correlated with the heights estimated earlier, and provide information on the packing of oligos: consistently with what reported earlier, in fact, the increased charge screening caused by the higher salt concentration, allows oligos to form denser monolayers; this effect is particularly evident for the HDM 0% for which a density increase of about 14% was observed, while a less pronounced variation was detected for the HDM 33% systems, where the higher salt concentration showed to yield a 7% more compact monolayer. It is furthermore

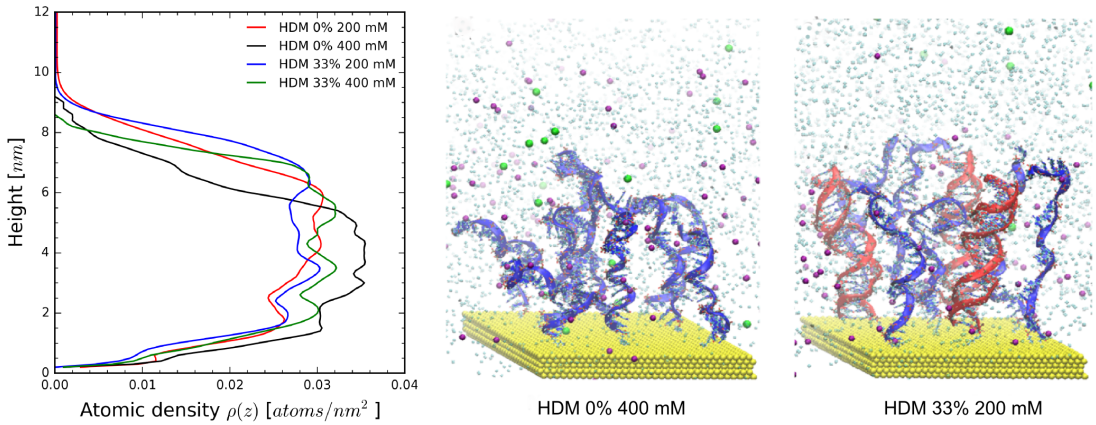


Figure 4.9: *Left* : atomic density $\rho(z)$ averaged over the trajectory as a function of the distance from the metal surface. *Right:* two snapshots taken from the trajectories displaying the arrangements of chains within the monolayers.

noticeable how the density for HDM 0% is higher than the one calculated for the HDM 33% system: this finding is quite surprising, since the total number of strands is higher in the second case and so was the expected density value. The two snapshots on the right hand side of fig. 4.9 give an intuitive picture of the configuration of oligos that results from densification: at 200mM of NaCl, ssDNA strands assume a stretched configuration and align to the more rigid dsDNA (rendered in red); on the contrary, at 400 mM, ssDNAs assume a coiled configurations which are responsible for the observed compaction effect.

The pivoting about the grafting site for bases positioned farther from the metal surface yields larger lateral fluctuations that determine the monotonic decrease of $\rho(z)$ for higher z values. The slope of the density curves, in fact, gives additional information of the morphology and the stiffness of the monolayers: for partially hybridised HDMs, the density decreases abruptly, suggesting that the sparse dsDNAs are sufficient to confer an increased rigidity to the monolayer, consistently with what has been observed so far; on the contrary, for HDM 0% patches, the slope is gradual, indicating that the motion of single stranded chains is less restrained. This, in turn, provides information on the HDM surface roughness: a flatter top of the density curve corresponds to a smoother and more uniform patch, while a gentle slope is associated with a rougher surface. We speculate that this information might be useful for the explanation of the different effectiveness observed in enzymatic reactions taking place on surface-bound DNA nanostructures [186, 187, 162].

The general agreement with AFM data suggests that the HDM molecular model devised for this study is sound and validates the energy analysis presented earlier. It is worth noting at this point that while AFM heights measured are representative of a value which is averaged

over the cantilever's tip, whose effective area is about 80 nm^2. In this respect, the simulations conducted in this work have the limitation of representing just a small portion of a micrometric-sized monolayer; on the other hand they provide an atomistic resolution which is otherwise not achievable with any other experimental techniques. Another observation that is worth doing concerns the simulation time, which was limited to 100 ns by the size of the system (more than 100k atoms). Longer integration times would be helpful to fully characterise the complex conformational space of the monolayer, while the model presented herein is focused on providing a simplified yet physically meaningful representation for final hybridisation states of DNA monolayers and the effect of a different ionic concentration, for which the sampling was contained compatibly with the available computational resources.

4.4 Coarse-grained model of HDMs

In order to overcome the limitations describe earlier, we intended to extend this study by employing a recently refined coarse-grained potential for DNA. The 3SPN.2 potential [188] was parametrised to match experimental melting curves of nucleic acids and demonstrated to describe successfully the kinetics and the thermodynamics of hybridisation of oligos in solution. In this model, each nucleotide is represented by three particles only, while a specific classical potential energy defines the contributions pertaining to intra- and inter- molecular interactions; in particular, stacking interaction and h-bonds are optimised to reproduce the stability of both double and single DNA strands. The number of degree of freedom is, thus, significantly reduced, while a sound description of the system is maintained. The adoption of this potential for HDM systems was considered as a promising strategy to extend simulations to the microsecond timescale, for system sizes larger bigger than the AFM tip ($\sim$400 nm^2). In particular, the idea was:

- to study surface hybridisation process for various HDM densities, including the adsorption of ssDNA into the monolayer;
- to derive height/loading curves for different hybridisation fractions to simulate with accuracy AFM experiments;
- to study the effect of cholesteric interaction in the packing of oligos.

Preliminary coarse-grained HDM systems were created, using the information extracted from MD trajectories to fit missing terms such as the presence of the aliphatic linker. The first simulations have been carried out with LAMMPS packages, for which 3SPN.2 potential has been

implemented. However, a bug in the MPI implementation of the code resulted in segmentation errors when used in conjunction with periodic boundary conditions. Various attempts have been made to tackle this issue, both on our part and by the De Pablo group at Chicago University, where originally the model was developed, without producing any significant improvement. For this reason, since the serial execution of the code did not show to provide the desired speedup of calculations, the development of the research in this direction was forcedly dismissed.

4.5 Summary

In this chapter, MD simulations and AFM experiments were used to investigate DNA high-density monolayers, which constitute important technologic platforms for the functionalisation of nano-devices. We focused on the characterisation of HDMs in the context of surface hybridisation, in order to establish how many ssDNAs are able to pair with the complementary sequences for a density of grafted oligos close to the observed experimental limit. In order to do so, MD simulations have been carried out on a series of molecular systems of varying hybridisation fractions, representing the "initial" (HDM 0%) and different possible "final" hybridised states, that are expected to be formed in the hypothesis of long reaction times. Generally, a good consistency between computational and experimental findings was observed. The results, in fact, confirm the existence of the hybridization limit, consistently with the evidences reported by other works. In particular, for a HDM of having density of about 1.2×10^{13} $strands/cm^2$, no more than 44% of the ssDNAs are expected to successfully hybridise, while 33% is the predicted most stable fraction: this is explained by an energy penalty associated with the enhanced electrostatic repulsion between negatively charged oligos which is observed above this threshold. Simulated and measured height profiles also revealed that the HDM structures behave rather differently for an increasing hybridization, becoming progressively less dense and sensible to the ion concentration; the latter was observed to induce more compact structures for un-hybridised monolayers. These findings evidence the role of ion screening and electrostatic repulsions between DNA strands in determining the packing order of oligos and affecting profoundly the monolayers' morphology.

The use of coarse grained potential for nucleic acids could be used in the framework outlined by this work to extend the study to monolayers of larger size and/or for longer simulation times at different surface densities. Some work has been carried on in this direction, in particularly to model realistically AFM height/load curves and, more recently, to investigate the influence of partially matching dsDNA sequences. However, it was not possible to carry out most of the proposed investigations for some technical difficulties that were partially solved just in the last

few months.

Chapter 5

Outlook and conclusions

Shedding light on the interaction between molecules at the atomistic level is a fundamental part of the process of designing functional self-assembled platforms that can respond to precise technological needs. We have reported here on the work carried out within this framework about three distinct molecular systems which have been studied through the development and application of a series of computational models which combined the use of classical potentials and DFT methods: in this way, experimental evidences have been supported with quantitative results aimed at providing insightful information on the aggregation of polyaromatic gelators, polycyclic aromatic hydrocarbons and nucleic acids.

In the first investigation, the study of the interactions established between a particular type of poly-aromatic binaphtol derivatives have been carried out with the purpose of providing a rationale for the observed sol-gel transition, with particular focus on trying to highlight its dependence on the chiral activity of the solution and to assess the effect of small modifications of the native species, such as methylation. We have shown that these molecules have two main conformations of similar energy that can undergo interconversion upon the rotation of one of the lateral groups around the central benzene. The energy barrier calculated for this conformational transition indicates the conversion mechanism is expected to take place at room temperature: as consequence both atropoisomers are likely to be found in the solution in nearly equal proportion. However, the two structures have a different dipole moment which is expected to influence their solubility in moderately polar solvents, as confirmed by the solvation energy calculated in methanol. The most relevant binding configurations between pairs of gelators have been detected by sampling the conformational space via MD. This allowed us to show that cis-conformers are able to interact more strongly with respect to all other pairs, while the analogous heterochiral pairs showed a significant decreased binding energy, which is likely attributed to

the diminished number of favourable configurations produced by steric hindrance.

Extensive simulations carried out on systems containing multiple molecules solvated in methanol have not revealed significant aggregation patterns. However, the formation of two interesting dimers has been observed, both characterised by the establishment of a series of $\pi - \pi$ and $ch - \pi$ interactions between the aromatic rings and hydrogen atoms. Importantly, these dimers were not observed to have comparable energies when formed by heterochiral atropoisomers. This observation has led us to consider the hypothesis that gelation could be explained by the aggregation of dimers. In this direction, both the energy of clusters formed by 10 dimers of different chiral compositions and their subsequent aggregation modelled with a fcc lattice provided some evidence for the existence of a energy difference favouring the aggregation of homochiral dimers against heterochiral one. This point was further addressed by the sampling of the interaction between pair of dimers through a Monte Carlo algorithm which revealed that, while the energy distributions do no display significant differences able to support the formation of a gel, the existence of a set of repetitive binding configurations for homochiral dimers constitutes a possible directional aggregation pattern. This hypothesis was tested with a simple Monte Carlo lattice model, showing that the stability of the aggregate varies substantially with the enantiomeric excess, even for small difference of the pair interaction energies.

The results outlined in this first chapter, provide substantial elements to further the study towards a more thorough understanding of the relation between the binding configurations and the resulting aggregation phenomenon. In this framework, the lattice Monte Carlo method could be extended to account for a variable number of particle in a three dimensional lattice, with the explicit addition of the contribution of solvation energy. This approach could, in particular, benefit from deep learning algorithm, able to map the non-linearities of the energies profile as a function of the orientation of between two dimers.

In the second study, part of the research work conducted in collaboration with BP has been presented. The binding energies of several PAH dimers were calculated via DFT and compared against the adsorption energies of the same PAH molecules onto a graphene plate. These energies once divided for the number of carbon atoms evidenced that smaller PAHs are capable of binding more strongly to graphene than they do to each other. The opposite energy trends constitute a particularly interesting finding that can be exploited for the design of molecules able to bind efficiently to complex aromatic substrates, which are known to form, in hydrocarbon mixtures during operative conditions in presence of reactive radicals. In this direction, a dendritic molecule was devised to test the effect of multiple small aromatic functionalisations at sticking onto highly defective aromatic substrates, thus enabling the formation of a hydrocarbon layer onto otherwise insoluble particles. The proposed mechanism has been tested through MD

simulations for three different model surfaces of decreasing aromatic fractions, solvated with long hydrocarbon molecules. In this context, the DFT energy profiles have been used to fit a corrective term for aromatic interaction for the classical potential. The results have evidenced the occurrence of multiple sticking events, that have been observed on both a flat graphitic plate and heterogeneous surfaces functionalised with methyl groups: importantly, smaller substituent not only interact more efficiently, but can also adapt to the varying degree of aromaticity of a real polyaromatic surfaces. Surprisingly, in fact, the small functionalities have been observed to bind more promptly to highly defective surfaces; the reason for this was found in the different cohesion of the first adlayer of hexadecane molecules, whose energy was observed to decrease significantly with the fraction of aromatic sites, thus favouring the penetration of the small phenyl substituents.

In the last work, MD simulations have being carried out to characterise DNA monolayers in the context of surface hybridisation. A portion of a DNA monolayer was modelled atomistically as to represent increasing hybridisation states in order to provide a rationale for the hybridisation limit evidenced by experiments in the high-density regime. The calculated energy profile as a function of the fraction of paired strands identified a minimum in correspondence to the 33% state, while no more that 44% of chains are expected to pair on average. This fraction corresponds to the maximum one for which double strands can be prevented from occupying first neighbouring positions, thus suggesting the role of the increasing electrostatic repulsion in hindering the hybridisation of additional single strand. The validation of results have been conducted through the comparison of AFM height measurements with the equilibrium strands height within the HDM, allowing also to assess the effect of ionic strength on the monolayer's morphology.

Possible development of this study includes the used of coarse-grained potentials that can be used to model the interaction of adsorbates with the monolayer and the surface hybridisation. The set up for these calculations has already been configured and some results have been recently obtained (article in preparation), while substantial bugs in the code limited the applicability of the research strategy, exploiting a reduce representation of DNAs for an enhanced sampling.

Together these three works represent a unique and original contribution to as many problems originated from experimental investigations. The picture emerging from the studies highlights the importance of aromatic interactions as key elements capable of controlling the aggregation of molecules. In particular, we observed that effectiveness of $\pi - \pi$ stacking and $ch - \pi$ bonds is closely related to the geometrical match between molecules and is found to be particularly important for systems characterised by moderate electrostatic contributions. More work needs to be done in this direction, and specifically, for the development of classical potentials that can

reproduce the directionality of aromatic interactions in a more accurate manner.

Bibliography

[1] T. D. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford Science Pubblications, London, 1987.

[2] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, Inc., New York, 2001.

[3] M. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford Graduate Texts, Oxford, 2001.

[4] B. J. Alder and T. E. Wainwright. Phase Transition for a Hard Sphere System. *J. Chem. Phys.*, 27(5):1208, 1957.

[5] A. Rahman. Correlations in the Motion of Atoms in Liquid Argon. *Phys. Rev.*, 136(2A):A405–A411, oct 1964.

[6] F. H. Stillinger, B. Laboratories, and M. Hill. Improved simulation of liquid water by molecular dynamics*. 60(4):1545–1557, 1974.

[7] J. A. McCammon. Dynamics of folded proteins. *Nature*, 267(5612):585–590, jun 1977.

[8] Gabor Csányi, T. Albaret, M. C. Payne, and A. De Vita. "Learn on the fly": A hybrid classical and quantum-mechanical molecular dynamics simulation. *Phys. Rev. Lett.*, 93(17):1–4, 2004.

[9] A. Aksimentiev and K. Schulten. Imaging alpha-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys. J.*, 88(6):3745–61, jun 2005.

[10] M. Orsi, W. E. Sanderson, and J. W. Essex. Permeability of small molecules through a lipid bilayer: a multiscale simulation study. *J. Phys. Chem. B*, 113(35):12019–29, sep 2009.

[11] E. Tajkhorshid, P. Nollert, and M. Jensen. Control of the Selectivity of the Aquaporin Water Channel Family by Global Orientational Tuning. 296(April), 2002.

[12] G. Doni, M. A. Kostiainen, A. Danani, and G. M. Pavan. Generation-dependent molecular recognition controls self-assembly in supramolecular dendron-virus complexes. *Nano Lett.*, 11(2):723–8, feb 2011.

[13] M. B. Ulmschneider, C. Bagnéris, E. C Mccusker, P. G. Decaen, and M. Delling. Molecular dynamics of ion transport through the open conformation of a bacterial voltage-gated sodium channel. pages 2–7, 2013.

[14] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25:1157–1174, 2004.

[15] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.

[16] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25(13):1656–76, oct 2004.

[17] W. L. Jorgensen and J. Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110(6):1657–1666, mar 1988.

[18] M. Levitt and S. Lifson. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.*, 46(2):269–279, 1969.

[19] L. Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159:98–103, 1967.

[20] P. P. Edwald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 54:253–287, 1921.

[21] N. Karasawa and W. A. Goddard III. Acceleration of Convergence for Lattice Sums. *J. Phys. Chem.*, 93(7902):7320–7327, 1989.

[22] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(1993):10089, 1993.

[23] B. Luty, I. G. Tironi, and W. F. van Gunsteren. Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *J. Chem. Phys.*, 103(1995):3014–3021, 1995.

[24] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(1980):2384–2393, 1980.

[25] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[26] S. A. Adelman and J. D. Doll. Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *J. Chem. Phys.*, 64:2375–2388, 1976.

[27] S. Nose. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1984):511–519, 1984.

[28] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.

[29] M. Parrinello and a. Rahman. Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.

[30] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101(5):4177, 1994.

[31] W. L. Jorgensen and J. Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U. S. A.*, 102(19):6665–6670, 2005.

[32] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926, 1983.

[33] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[34] M. M. Conde, M. a. Gonzalez, J. L F Abascal, and C. Vega. Determining the phase diagram of water from direct coexistence simulations: The phase diagram of the TIP4P/2005 model revisited. *J. Chem. Phys.*, 139(15), 2013.

[35] M. Yamada, S. Mossa, H. E. Stanley, and F. Sciortino. Interplay Between Time-Temperature-Transformation and the Liquid-Liquid Phase Transition in Water. page 4, 2002.

[36] Z. Yu, M. P. Jacobson, J. Josovitz, C. S. Rapp, and R. A. Friesner. First-Shell Solvation of Ion Pairs: Correction of Systematic Errors in Implicit Solvent Models. *J. Phys. Chem. B*, 108(21):6643–6654, 2004.

[37] M. Orsi. Comparative assessment of the ELBA coarse-grained model for water. *Mol. Phys.*, 112(11):1566–1576, oct 2013.

[38] D. Eisenberg and A. D. McLachlan. Solvation energy in protein folding and binding. *Nature*, 319(6050):199–203, 1986.

[39] J. Kleinjung, W. R. P. Scott, J. R. Allison, W. F. Van Gunsteren, and F. Fraternali. Implicit solvation parameters derived from explicit water forces in large-scale molecular dynamics simulations. *J. Chem. Theory Comput.*, 8(7):2391–2403, 2012.

[40] W. Hasel, T. F. Hendrickson, and W. C. Still. A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput. Methodol.*, 1(2):103–116, 1988.

[41] M Gerstein and Fm Richards. Protein geometry: volumes, areas, and distances. *Int. Tables . . .*, 318:1–28, 2001.

[42] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112(16):6127–6129, 1990.

[43] A. Onufriev, D. A. Case, and D. Bashford. Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.*, 23(14):1297–1304, 2002.

[44] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.*, 33(12):889–897, 2000.

[45] G. Rastelli, A. Del Rio, G. Degliesposti, and M. Sgobba. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.*, 31(4):797–810, 2010.

[46] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der Waals density functional for general geometries. *Phys. Rev. Lett.*, 92(24):246401–1, 2004.

[47] J. Klimes, D. R. Bowler, and A. Michaelides. Chemical accuracy for the van der Waals density functional. *J. Phys. Condens. Matter*, 22:022201, 2010.

[48] J. D. Pack and H. J. Monkhorst. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 16(4):1748–1749, 1977.

[49] R. P. Feynman. Forces in molecules. *Phys. Rev.*, 56(4):340–343, 1939.

[50] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B.. Robertson, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvary, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sangui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. AMBER 11, 2010.

[51] S. Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics., 1995.

[52] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeerschd, E. Zurek, and G. R. Hutchison. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.*, 4(8):1–17, 2012.

[53] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P Umari, and R. M. Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter*, 21(39):395502, 2009.

[54] G. Kresse and J. Furthmuller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, 1996.

[55] A. R. Hirst, B. Escuder, J. F. Miravet, and D. K. Smith. High-tech applications of self-assembling supramolecular nanostructured gel-phase materials: From regenerative medicine to electronic devices. *Angew. Chemie - Int. Ed.*, 47:8002–8018, 2008.

[56] D. J. Cornwell and D. K. Smith. Expanding the scope of gels- combining polymers with low-molecular-weight gelators to yield modified self-assembling smart materials with high-tech applications. *Mater. Horiz.*, 00:1–15, 2015.

[57] D. K. Smith. Lost in translation? Chirality effects in the self-assembly of nanostructured gel-phase materials. *Chem. Soc. Rev.*, 38(3):684–694, 2009.

[58] T. Naota and H. Koori. Molecules that assemble by sound: an application to the instant gelation of stable organic fluids. *J. Am. Chem. Soc.*, 127(26):9324–5, jul 2005.

[59] J. Wu, T. Yi, Q. Xia, Y. Zou, F. Liu, J. Dong, T. Shu, F. Li, and C. Huang. Tunable gel formation by both sonication and thermal processing in a cholesterol-based self-assembly system. *Chemistry*, 15(25):6234–43, jun 2009.

[60] D. Ke, C. Zhan, A. D. Q. Li, and J. Yao. Morphological Transformation between Nanofibers and Vesicles in a Controllable Bipyridine-Tripeptide Self-Assembly. *Angew. Chemie*, 123(16):3799–3803, apr 2011.

[61] G. Cravotto, S. Tabasso, D. Carnaroglio, and E. Calcio. Microwave, ultrasound and ball mill procedures for bio-waste valorisation. *Green Chem.*, nov 2014.

[62] A. R. Hirst, I. A. Coates, T. R. Boucheteau, J. F. Miravet, B. Escuder, V. Castelletto, I. W. Hamley, and D. K. Smith. Low-molecular-weight gelators: Elucidating the principles of gelation based on gelator solubility and a cooperative self-assembly model. *J. Am. Chem. Soc.*, 130:9113–9121, 2008.

[63] G. W. Brinker, C. J. and Scherer. *Sol-Gel Science*. Academic Press, Inc., 1990.

[64] J. H. van Esch. We can design molecular gelators, but do we understand them? *Langmuir*, 25(15):8392–4, aug 2009.

[65] D. Bardelang. Ultrasound induced gelation: a paradigm shift. *Soft Matter*, 5(10):1969, 2009.

[66] G. Cravotto and P. Cintas. Molecular self-assembly and patterning induced by sound waves. The case of gelation. *Chem. Soc. Rev.*, 38(9):2684–97, sep 2009.

[67] G. Cravotto, E. C. Gaudino, and P. Cintas. On the mechanochemical activation by ultrasound. *Chem. Soc. Rev.*, 42(18):7521–34, sep 2013.

[68] J. Ribas-Arino and D. Marx. Covalent Mechanochemistry : Theoretical Concepts and Computational Tools with Applications to Molecular Nanomechanics. 2012.

[69] S. D. Chakarova and E. Schröder. van der Waals interactions of polycyclic aromatic hydrocarbon dimers. *J. Chem. Phys.*, 122(5):54102, feb 2005.

[70] M. Caricato, A. Delforge, D. Bonifazi, D. Dondi, A. Mazzanti, and D. Pasini. Chiral nanostructuring of multivalent macrocycles in solution and on surfaces. *Org. Biomol. Chem.*, 13(12):3593–3601, 2015.

[71] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, 100(4):19824–19839, 1996.

[72] Ilja V. Khavrutskii and Anders Wallqvist. Computing relative free energies of solvation using single reference thermodynamic integration augmented with Hamiltonian Replica Exchange. *J. Chem. Theory Comput.*, 6(11):3427–3441, 2010.

[73] Sílvia Alexandra Martins, Sergio F. Sousa, Maria João Ramos, and Pedro Alexandrino Fernandes. Prediction of Solvation Free Energies with Thermodynamic Integration using the General Amber Force Field. *J. Chem. Theory Comput.*, 10:3570–3577, 2014.

[74] M. Orsi, W. Ding, and M. Palaiokostas. Direct mixing of atomistic solutes and coarse-grained water Direct mixing of atomistic solutes and coarse-grained water. *J. Chem. Theory Comput.*, 10:4684–4693, 2014.

[75] A. Aggeli, I .A. Nyrkova, M. Bell, R. Harding, L. Carrick, T. C. McLeish, A. N. Semenov, and N. Boden. Hierarchical self-assembly of chiral rod-like molecules as a model for peptide beta-sheet tapes, ribbons, fibrils, and fibers. *Proc. Natl. Acad. Sci. U. S. A.*, 98(21):11857–11862, 2001.

[76] H. Suezawa, S. Ishihara, Y. Umezawa, S. Tsuboyama, and M. Nishio. The aromatic CH-pi hydrogen bond as an important factor in determining the relative stability of diastereomeric salts relevant to enantiomeric resolution - A crystallographic database study. *European J. Org. Chem.*, pages 4816–4822, 2004.

[77] C. D. Zeinalipour-Yazdi and D. P. Pullman. Correlation of Polarizabilities with Van Der Waals Interactions in π -systems. 110(47):24260–24265, 2006.

[78] Z. Zhang and S. C. Glotzer. Self-assembly of patchy particles. *Nano Lett.*, 4(8):1407–1413, 2004.

[79] B. Ruan and A. M. Jacobi. Ultrasonication effects on thermal and rheological properties of carbon nanotube suspensions. *Nanoscale Res. Lett.*, 7(1):127, 2012.

[80] R. E. Miles. On Random Rotations in R^3 Miscellanea. 52(3):636–639, 1965.

[81] V. Tsui and D. A. Case. Theory and applications of the Generalized Born solvation model in macromolecular simulations. *Biopolymers*, 56(4):275–291, 2000.

[82] J. G. Gay and B. J. Berne. Modification of the overlap potential to mimic a linear site-site potential. *J. Chem. Phys.*, 74(1981):3316, 1981.

[83] A. Della Pia, M. Riello, A. Floris, D. Stassen, T. S. Jones, D. Bonifazi, A. De Vita, and G. Costantini. Anomalous coarsening driven by reversible charge transfer in supramolecular assembly at metal-organic interfaces. pages 12356–12364, 2014.

[84] A. Coniglio, H. E. Stanley, and W. Klein. Site-bond correlated-percolation problem: A statistical mechanical model of polymer gelation. *Phys. Rev. Lett.*, 42:518–522, 1979.

[85] S. George, S. Balla, V. Gautam, and M. Gautam. Effect of diesel soot on lubricant oil viscosity. *Tribol. Int.*, 40:809–818, 2007.

[86] S. George, S. Balla, and M. Gautam. Effect of diesel soot contaminated oil on engine wear. *Wear*, 262(September 2006):1113–1122, 2007.

[87] J. O. Müller, D. S. Su, U. Wild, and R. Schlögl. Bulk and surface structural investigations of diesel engine soot and carbon black. *Phys. Chem. Chem. Phys.*, 9:4018–4025, 2007.

[88] P. Oleszczuk. Persistence of polycyclic aromatic hydrocarbons (PAHs) in sewage sludge-amended soil. *Chemosphere*, 65(9):1616–26, nov 2006.

[89] C. E. Cerniglia. Biodegradation of polycyclic aromatic hydrocarbons. *Curr. Opin. Biotechnol.*, 4(3):331–338, jun 1993.

[90] S. K. Samanta, O. V. Singh, and R. K. Jain. Polycyclic aromatic hydrocarbons: environmental pollution and bioremediation. *Trends Biotechnol.*, 20(6):243–248, jun 2002.

[91] S. M. Bamforth and I. Singleton. Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions. *J. Chem. Technol. Biotechnol.*, 80(7):723–736, 2005.

[92] K. Srogi. Monitoring of environmental exposure to polycyclic aromatic hydrocarbons: a review. *Environ. Chem. Lett.*, 5(4):169–195, mar 2007.

[93] C. E. Boström, P. Gerde, A. Hanberg, B. Jernström, C. Johansson, T. Kyrklund, A. Rannug, M. Törnqvist, K. Victorin, and R. Westerholm. Cancer Risk Assessment, Indicators, and Guidelines for Polycyclic Aromatic Hydrocarbons in the Ambient Air. *Environ. Health Perspect.*, 110(s3):451–489, feb 2002.

[94] B. D. Jeffy, R. B. Chirnomas, and D. F. Romagnolo. Epigenetics of breast cancer: polycyclic aromatic hydrocarbons as risk factors. *Environ. Mol. Mutagen.*, 39(2-3):235–44, jan 2002.

[95] L. E. Smith, M. F. Denissenko, W. P. Bennett, H. Li, S. Amin, M. S. Tang, and G. P. Pfeifer. Targeting of Lung Cancer Mutational Hotspots by Polycyclic Aromatic Hydrocarbons. *JNCI J. Natl. Cancer Inst.*, 92(10):803–811, may 2000.

[96] A. K. Geim and K. S. Novoselov. The rise of graphene. *Nat. Mater.*, 6(3):183–91, mar 2007.

[97] A. K. Geim. Graphene: status and prospects. *Science*, 324(5934):1530–4, jun 2009.

[98] D. Quiñonero, C. Garau, A. Frontera, P. Ballester, A. Costa, and P. M. Deyà. Structure and binding energy of anion-π and cation-π complexes: A comparison of MP2, RI-MP2, DFT, and DF-DFT methods. *J. Phys. Chem. A*, 109:4632–4637, 2005.

[99] H. Lang, R. Packheiser, and B. Walfort. Organometallic pi-tweezers, NCN pincers, and ferrocenes as molecular "Tinkertoys" in the synthesis of multiheterometallic transition-metal complexes. *Organometallics*, 25(8):1836–1850, 2006.

[100] A. Puzder, M. Dion, and D. C. Langreth. Binding energies in benzene dimers: Nonlocal density functional calculations. *J. Chem. Phys.*, 124:1–17, 2006.

[101] O. Bludský, M. Rubeš, P. Soldán, and P. Nachtigall. Investigation of the benzene-dimer potential energy surface: DFT/CCSD(T) correction scheme. *J. Chem. Phys.*, 128, 2008.

[102] I. D. Mackie and G. A. Dilabio. Accurate dispersion interactions from standard density-functional theory methods with small basis sets. *Phys. Chem. Chem. Phys.*, 12(d):6092–6098, 2010.

[103] S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27(15):1787–1799, 2006.

[104] S. D. Chakarova-Käck, A. Vojvodic, J. Kleis, P. Hyldgaard, and E. Schröder. Binding of polycyclic aromatic hydrocarbons and graphene dimers in density functional theory. *New J. Phys.*, 12(1):013017, jan 2010.

[105] S. M. Kozlov, F. Viñes, and A. Görling. On the interaction of polycyclic aromatic compounds with graphene. *Carbon N. Y.*, 50:2482–2492, 2012.

[106] R. Zacharia, H. Ulbricht, and T. Hertel. Interlayer cohesive energy of graphite from thermal desorption of polyaromatic hydrocarbons. *Phys. Rev. B - Condens. Matter Mater. Phys.*, 69:1–7, 2004.

[107] S. D. Chakarova-Käck, E. Schröder, B. I. Lundqvist, and D. C. Langreth. Application of van der Waals density functional to an extended system: Adsorption of benzene and naphthalene on graphite. *Phys. Rev. Lett.*, 96(April):1–4, 2006.

[108] C. Feng, C. S. Lin, W. Fan, R. Q. Zhang, and M. A. Van Hove. Stacking of polycyclic aromatic hydrocarbons as prototype for graphene multilayers, studied using density functional theory augmented with a dispersion term. *J. Chem. Phys.*, 131(2009), 2009.

[109] A. Gulans, M. J. Puska, and R. M. Nieminen. Linear-scaling self-consistent implementation of the van der Waals density functional. *Phys. Rev. B - Condens. Matter Mater. Phys.*, 79(20):1–4, 2009.

[110] X. Chen, F. Tian, C. Persson, W. Duan, and N. Chen. Interlayer interactions in graphites. *Sci. Rep.*, 3:3046, 2013.

[111] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, may 1995.

[112] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B*, 105:6474–6487, 2001.

[113] E. R. Zubarev, M. U. Pralle, E. D. Sone, and S. I. Stupp. Self-Assembly of Dendron Rodcoil Molecules into Nanoribbons. *J. Am. Chem. Soc.*, 123(17):4105–4106, 2001.

[114] G. Ungar, Y. Liu, X. Zeng, V. Percec, and W. D. Cho. Giant supramolecular liquid crystal lattice. *Science*, 299(5610):1208–11, feb 2003.

[115] B. M. Rosen, C. J. Wilson, D. A. Wilson, M. Peterca, M. R. Imam, and V. Percec. Dendron-mediated self-assembly, disassembly, and self-organization of complex systems. *Chem. Rev.*, 109(11):6275–540, nov 2009.

[116] M. Garzoni, N. Cheval, A. Fahmi, A. Danani, and G. M. Pavan. Ion-selective controlled assembly of dendrimer-based functional nanofibers and their ionic-competitive disassembly. *J. Am. Chem. Soc.*, 134(7):3349–57, feb 2012.

[117] M. A. Herrero and F. M. Toma. Synthesis and characterization of a carbon nanotube-dendron series for efficient siRNA delivery. *J. Am. Chem. Soc.*, 131(28):9843–8, jul 2009.

[118] W. She, K. Luo, C. Zhang, G. Wang, Y. Geng, L. Li, B. He, and Z. Gu. The potential of self-assembled, pH-responsive nanoparticles of mPEGylated peptide dendron-doxorubicin conjugates for cancer therapy. *Biomaterials*, 34(5):1613–23, feb 2013.

[119] S. M. Bromfield, P. Posocco, M. Fermeglia, S. Pricl, J. Rodríguez-López, and D. K. Smith. A simple new competition assay for heparin binding in serum applied to multivalent PAMAM dendrimers. *Chem. Commun. (Camb).*, 49:4830–2, 2013.

[120] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.*, 21(2):132–146, 2000.

[121] L. Liao, Z. Song, Y. Zhou, H. Wang, Q. Xie, H. Peng, and Z. Liu. Photoinduced methylation of graphene. *Small*, 9(8):1348–1352, 2013.

[122] F. Liang, M. Lu, T. C. Keener, Z. Liu, and S. J. Khang. The organic composition of diesel particulate matter, diesel fuel and engine oil of a non-road diesel generator. *J. Environ. Monit.*, 7:983–988, 2005.

[123] J. H. Dymond, K. J. Young, and J. D. Isdale. Transport properties of nonelectrolyte liquid mixtures-II. Viscosity coefficients for the n-hexane + n-hexadecane system at temperatures from 25 to 100 C at pressures up to the freezing pressure or 500 MPa. *Int. J. Thermophys.*, 1(4):345–373, 1980.

[124] S. W. I. Siu, K. Pluhackova, and R. A. Böckmann. Optimization of the OPLS-AA force field for long hydrocarbons. *J. Chem. Theory Comput.*, 8:1459–1470, 2012.

[125] B. Chen, M. G. Martin, and J. I. Siepmann. Thermodynamic properties of the williams, OPLS-AA, and MMFF94 all-atom force fields for normal alkanes. *J. Phys. Chem. B*, 102(98):2578–2586, 1998.

[126] M. G. Martin and J. I. Siepmann. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n -Alkanes. *J. Phys. Chem. B*, 102(97):2569–2577, 1998.

[127] D. K. Dysthe, A. H. Fuchs, and B. Rousseau. Fluid transport properties by equilibrium molecular dynamics. III. Evaluation of united atom interaction potential models for pure alkanes. *J. Chem. Phys.*, 112(2000):7581–7590, 2000.

[128] M. J. Penna, M. Mijajlovic, and M. J. Biggs. Molecular-level understanding of protein adsorption at the interface between water and a strongly interacting uncharged solid surface. *J. Am. Chem. Soc.*, 136(14):5323–5331, 2014.

[129] D. P. Smith, H. Hörber, C. Gerber, and G. Binnig. Smectic liquid crystal monolayers on graphite observed by scanning tunneling microscopy. *Science*, 245(4913):43–45, 1989.

[130] K. Uosaki and R. Yamada. Formation of two-dimensional crystals of alkanes on the Au(111) surface in neat liquid. *J. Am. Chem. Soc.*, 121(16):4090–4091, 1999.

[131] T. Yang, S. Berber, J. F. Liu, G. P. Miller, and D. Tománek. Self-assembly of long chain alkanes and their derivatives on graphite. *J. Chem. Phys.*, 128(12), 2008.

[132] J. S. Yang, C. L. Yang, M. S. Wang, B. D. Chen, and X. G. Ma. Crystallization of alkane melts induced by carbon nanotubes and graphene nanosheets: a molecular dynamics simulation study. *Phys. Chem. Chem. Phys.*, 13(34):15476–15482, 2011.

[133] T. Junno, K. Deppert, L. Montelius, and L. Samuelson. Controlled manipulation of nanoparticles with an atomic force microscope. *Appl. Phys. Lett.*, 66(26):3627, 1995.

[134] D. M. Eigler and E. K. Schweizer. Positioning single atoms with a scanning tunnelling microscope. *Nature*, 344(6266):524–526, 1990.

[135] P. W. K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.

[136] C. M. Niemeyer. Semisynthetic DNA-protein conjugates for biosensing and nanofabrication. *Angew. Chem. Int. Ed. Engl.*, 49(7):1200–16, 2010.

[137] T. Vo-Dinh. Development of a DNA biochip : principle and applications 1. *Sensors Actuators B*, 51:52–59, 2008.

[138] S. Nimse, K. Song, M. Sonawane, D. Sayyed, and T. Kim. Immobilization Techniques for Microarray: Challenges and Applications. *Sensors*, 14(12):22208–22229, 2014.

[139] M. L. Phelan and S. Nock. Generation of bioreagents for protein chips. *Proteomics*, 3(11):2123–34, nov 2003.

[140] C. Boozer, J. Ladd, S. Chen, and S. Jiang. DNA-directed protein immobilization for simultaneous detection of multiple analytes by surface plasmon resonance biosensor. *Anal. Chem.*, 78(5):1515–9, mar 2006.

[141] R. Levicky, T. M. Herne, M. J. Tarlov, and S. K. Satija. Using self-assembly to control the structure of DNA monolayers on gold: A neutron reflectivity study. *J. Am. Chem. Soc.*, 120(6):9787–9792, 1998.

[142] R. Maoz, S. R. Cohen, and J. Sagiv. Nanoelectrochemical Patterning of Monolayer Surfaces: Toward Spatially Defined Self-Assembly of Nanostructures. *Adv. Mater.*, 11(1):55–61, 1999.

[143] W. Cheng, M. J. Campolongo, J. J. Cha, S. J. Tan, C. C. Umbach, D. A. Muller, and D. Luo. Free-standing nanoparticle superlattice sheets controlled by DNA. *Nat. Mater.*, 8(6):519–525, 2009.

[144] M. Melli, G. Scoles, and M. Lazzarino. Fast detection of biomolecules in diffusion-limited regime using micromechanical pillars. *ACS Nano*, 5(10):7928–35, 2011.

[145] Y. Ahn, J. K. Saha, G. C. Schatz, and J. Jang. Molecular Dynamics Study of the Formation of a Self-Assembled Monolayer on Gold. *J. Phys. Chem. C*, 115(21):10668–10674, 2011.

[146] S. Xu, S. Miller, P. E. Laibinis, and G. Liu. Fabrication of Nanometer Scale Patterns within Self-Assembled Monolayers by Nanografting. *Langmuir*, 15(21):7244–7251, 1999.

[147] E. Mirmomtaz, M. Castronovo, C. Grunwald, F. Bano, D. Scaini, A. A. Ensafi, G. Scoles, and L. Casalis. Quantitative study of the effect of coverage on the hybridization efficiency of surface-bound DNA nanostructures. *Nano Lett.*, 8(12):4134–9, 2008.

[148] S. Ryu and G. C. Schatz. Nanografting: Modeling and simulation. *J. Am. Chem. Soc.*, 128(11):11563–11573, 2006.

[149] J. L. Arlett, E. B. Myers, and M. L. Roukes. Comparative advantages of mechanical biosensors. *Nat. Nanotechnol.*, 6(4):203–15, 2011.

[150] F. G. A Faas, B. Rieger, L. J. Van Vliet, and D. I. Cherny. DNA deformations near charged surfaces: Electron and atomic force microscopy views. *Biophys. J.*, 97(4):1148–1157, 2009.

[151] A. B. Ozel. *Theoretical and Experimental Investigation of the Impact of Surfaces on DNA Melting Temperature.* ProQuest, 2009.

[152] A. W. Peterson, R. J. Heaton, and R. M. Georgiadis. The effect of surface probe density on DNA hybridization. *Nucleic Acids Res.*, 29(24):5163–8, 2001.

[153] M. F. Hagan and A. K. Chakraborty. Hybridization dynamics of surface immobilized DNA. *J. Chem. Phys.*, 120(10):4958–68, 2004.

[154] I. Y. Wong and N. A. Melosh. An electrostatic model for DNA surface hybridization. *Biophys. J.*, 98(12):2954–63, 2010.

[155] J. Fritz, E. B. Cooper, S. Gaudet, P. K. Sorger, and S. R. Manalis. Electronic detection of DNA by its intrinsic molecular charge. *Proc. Natl. Acad. Sci. U. S. A.*, 99(22):14142–14146, 2002.

[156] C. Maffeo, R. Schöpflin, H. Brutzer, R. Stehr, A. Aksimentiev, G. Wedemann, and R. Seidel. DNA-DNA Interactions in Tight Supercoils Are Described by a Small Effective Charge Density. *Phys. Rev. Lett.*, 105(15):158101, oct 2010.

[157] G. S. Manning. Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions. III. An Analysis Based on the Mayer Ionic Solution Theory. *J. Chem. Phys.*, 51(1969):3249, 1969.

[158] F. Bano, L. Fruk, B. Sanavio, M. Glettenberg, L. Casalis, C. M. Niemeyer, and G. Scoles. Toward multiprotein nanoarrays using nanografting and DNA directed immobilization of proteins. *Nano Lett.*, 9(7):2614–8, 2009.

[159] B. Tinland, A. Pluen, J. Sturm, and G. Weill. Persistence Length of Single-Stranded DNA. *Macromolecules*, (30):5763–5765, 1997.

[160] Y. Lu, B. Weers, and N. C. Stellwagen. DNA persistence length revisited. *Biopolymers*, 61(4):261–75, 2002.

[161] A. Bosco, F. Bano, P. Parisse, L. Casalis, A. DeSimone, and C. Micheletti. Hybridization in nanostructured DNA monolayers probed by AFM: theory versus experiment. *Nanoscale*, 4(5):1734–41, 2012.

[162] M. Castronovo, A. Lucesoli, P. Parisse, A. Kurnikova, A. Malhotra, M. Grassi, G. Grassi, B. Scaggiante, L. Casalis, and G. Scoles. Two-dimensional enzyme diffusion in laterally confined DNA monolayers. *Nat. Commun.*, 2(may):297, 2011.

[163] A. Singh, S. Snyder, L. Lee, A. P. R. Johnston, F. Caruso, and Y. G. Yingling. Effect of oligonucleotide length on the assembly of DNA materials: molecular dynamics simulations of layer-by-layer DNA films. *Langmuir*, 26(22):17339–47, 2010.

[164] D. Poland and H. A. Scheraga. Occurrence of a phase transition in nucleic acid models. *J. Chem. Phys.*, 45(5):1464–1469, 1966.

[165] A. Jayaraman, C. K. Hall, and J. Genzer. Computer simulation study of probe-target hybridization in model DNA microarrays: effect of probe surface density and target concentration. *J. Chem. Phys.*, 127(14):144912, 2007.

[166] E J Sambriski, D C Schwartz, and J J de Pablo. Uncovering pathways in DNA oligonucleotide hybridization via transition state analysis. *Proc. Natl. Acad. Sci. U. S. A.*, 106(43):18125–30, oct 2009.

[167] T. E. Ouldridge, P. Šulc, F. Romano, J. P K Doye, and A. A. Louis. DNA hybridization kinetics: Zippering, internal displacement and sequence dependence. *Nucleic Acids Res.*, 41(19):8886–8895, 2013.

[168] Liqun Yao, Jim Sullivan, Jason Hower, Yi He, and Shaoyi Jiang. Packing structures of single-stranded DNA and double-stranded DNA thiolates on Au(111): a molecular simulation study. *J. Chem. Phys.*, 127(19):195101, nov 2007.

[169] A. V. Pinheiro, J. Nangreave, S. Jiang, H. Yan, and Y. Liu. Steric crowding and the kinetics of DNA hybridization within a DNA nanostructure system. *ACS Nano*, 6(6):5521–5530, 2012.

[170] B. Rai, P. Sathish, C. P. Malhotra, and K. G. Ayappa. Molecular Dynamic Simulations of Self-Assembled Alkylthiolate Monolayers on an Au (111) Surface. 98(111):3138–3144, 2004.

[171] J. K. Saha, Y. Ahn, H. Kim, G. C. Schatz, and J. Jang. Small Size Limit to Self-Assembled Monolayer Formation on Gold(111). *J. Phys. Chem. C*, 115(27):13193–13199, 2011.

[172] S. A. Harris, Z. A. Sands, and C. A. Laughton. Molecular dynamics simulations of duplex stretching reveal the importance of entropy in determining the biomechanical properties of DNA. *Biophys. J.*, 88(3):1684–91, 2005.

[173] O. S. Lee and G. C. Schatz. Interaction between DNAs on a gold surface. *J. Phys. Chem. C*, 113(36):15941–15947, 2009.

[174] One-Sun Lee and George C. Schatz. Molecular Dynamics Simulation of DNA-Functionalized Gold Nanoparticles. *J. Phys. Chem. C*, 113(6):2316–2321, 2009.

[175] D. Marenduzzo, E. Orlandini, A. Stasiak, D. W. Sumners, L. Tubiana, and C. Micheletti. DNA-DNA interactions in bacteriophage capsids are responsible for the observed DNA knotting. *Proc. Natl. Acad. Sci. U. S. A.*, 106(52):22269–22274, 2009.

[176] S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case. An All Atom Force Field for Simulation of Proteins and Nucleic Acids. *J. Comp. Chem.*, 7(2):230–252, 1986.

[177] M. Orozco, A. Pérez, A. Noy, and F. J. Luque. Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.*, 32(6):350–364, 2003.

[178] M. Zgarbová, M. Otyepka, J. Šponer, A. Mládek, P. Banáš, T. E. Cheatham, and P. Jurečka. Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, 7(9):2886–2902, 2011.

[179] J. P. Ryckaert, G. Ciccotti, and H. J.C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23(3):327–341, 1977.

[180] D. Peled, R. Naaman, and S. S. Daube. Packed DNA denatures on gold nanoparticles. *J. Phys. Chem. B*, 114(25):8581–8584, 2010.

[181] D. Peled, S. S. Daube, and R. Naaman. Selective enzymatic labeling to detect packing-induced denaturation of double-stranded DNA at interfaces. *Langmuir*, 24(20):11842–11846, 2008.

[182] P. G. de Gennes. Conformations of Polymers Attached to an Interface. *Macromolecules*, 13(19):1069–1075, 1980.

[183] S. Milner, T. Witten, and M. Cates. Theory of the grafted polymer brush. *Macromolecules*, 21(8):2610–2619, 1988.

[184] G. Doni, M. D. Nkoua Ngavouka, A. Barducci, P. Parisse, A. De Vita, G. Scoles, L. Casalis, and G. M Pavan. Structural and energetic basis for hybridization limits in high-density DNA monolayers. *Nanoscale*, 5(20):9988–93, 2013.

[185] P. S. Crozier and M. J. Stevens. Simulations of single grafted polyelectrolyte chains: ssDNA and dsDNA. *J. Chem. Phys.*, 118(8):3855–3860, 2003.

[186] Sung Yong Park, Abigail K R Lytton-Jean, Byeongdu Lee, Steven Weigand, George C Schatz, and Chad a Mirkin. DNA-programmable nanoparticle crystallization. *Nature*, 451(7178):553–6, jan 2008.

[187] P. Parisse, A. Vindigni, G. Scoles, and L. Casalis. In vitro enzyme comparative kinetics: Unwinding of surface-bound DNA nanostructures by RecQ and RecQ1. *J. Phys. Chem. Lett.*, 3(23):3532–3537, 2012.

[188] D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. De Pablo. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.*, 139(14), 2013.